The roles of visual expertise and visual input in the face inversion effect: Behavioral and neurocomputational evidence

Joseph P. McCleery [a], Lingyun Zhang [b], Liezhong Ge [c,*], Zhe Wang [c], Eric M. Christiansen [d], Kang Lee [e], Garrison W. Cottrell [b,*]

[a] *University of California, San Diego, Department of Psychology, 9500 Gilman Drive, Mail Code 0109, La Jolla, CA 92093-0109, United States*
[b] *University of California, San Diego, Department of Computer Science and Engineering, 9500 Gilman Drive, UCSD CSE 0404, La Jolla, CA 92093-0404, United States*
[c] *Zhejiang Sci-Tech University, Higher Education Park in Xiasha, Hangzhou, Zhejiang Province 310018, People's Republic of China*
[d] *Swarthmore College, 500 College Avenue, Swarthmore, PA 19081, United States*
[e] *Institute of Child Study, University of Toronto, 45 Walmer Road, Toronto, Ontario, Canada M5R 2X2*

Abstract

Research has shown that inverting faces significantly disrupts the processing of configural information, leading to a *face inversion effect*. We recently used a contextual priming technique to show that the presence or absence of the face inversion effect can be determined via the top–down activation of face versus non-face processing systems [Ge, L., Wang, Z., McCleery, J., & Lee, K. (2006). Activation of face expertise and the inversion effect. *Psychological Science, 17*(1), 12–16]. In the current study, we replicate these findings using the same technique but under different conditions. We then extend these findings through the application of a neural network model of face and Chinese character expertise systems. Results provide support for the hypothesis that a specialized face expertise system develops through extensive training of the visual system with upright faces, and that top–down mechanisms are capable of influencing when this face expertise system is engaged.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Face processing; Chinese characters; Neural network models; Neurocomputational models; Visual expertise; Visual discrimination; Configural processing; Featural processing

1. Introduction

Research has shown that inverting faces significantly disrupts the processing of configural information, leading to a *face inversion effect* (e.g., Searcy & Bartlett, 1996). The psychophysical result of this disruption of configural information is that inverting faces impairs discrimination performance more than does inverting non-face objects. Additionally, inverting faces during feature-based discriminations does not impair performance, but inverting faces

during a configural discrimination task does (e.g., Farah, Tanaka, & Drain, 1995; Freire, Lee, & Symons, 2000, but see Yovel & Kanwisher, 2004). Therefore, the evidence to date suggests that face processing involves a specialized type of configural processing, which may be associated with high levels of visual discrimination expertise (see e.g., Gauthier & Tarr, 1997). However, little is known about the direct relationships of visual expertise, configural processing, and the face inversion effect.

We recently found psychophysical evidence that the presence or absence of the face inversion effect can be determined via the top–down activation of face versus non-face processing systems (Ge et al., 2006). In this previous experiment, we primed Chinese participants with either

---

[*] Corresponding authors. Fax: +1 858 534 7029.
   E-mail addresses: glzh@psyzjus.cn (L. Ge), gary@ucsd.edu (G.W. Cottrell).

two face tasks (Real Face Discrimination Task followed by Schematic Face Discrimination Task) or two Chinese character tasks (Real Character Discrimination Task followed by Nonsense Character Discrimination Task) and then tested them on Ambiguous Figures that could be perceived as either faces or Chinese characters (see Fig. 1). In the study, all tasks involved exclusively configural discriminations. The reason we chose Chinese characters as comparison stimuli is that they are highly similar to faces on a number of dimensions, except for the fact that configural information plays an important role in face processing but not in character processing (see Table 1).

In our previous study, we found that participants in the face priming condition showed an inversion effect in both the face priming tasks as well as the Ambiguous Figures Task, whereas the participants in the character priming condition did not show an inversion effect in either the character priming tasks or the Ambiguous Figures Task
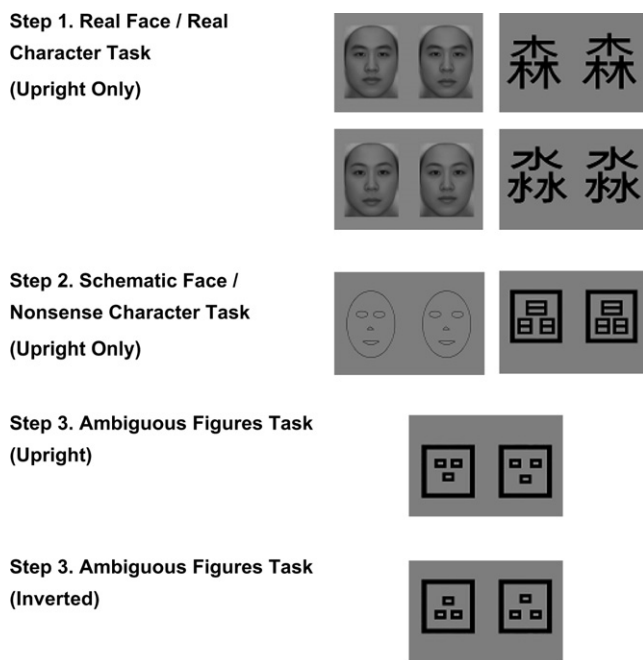

Fig. 1. Design of the Face and Character Conditions in Experiments 1a and 1b.

Table 1
Main similarities and differences between faces and Chinese characters

	Face	Chinese character
1. Extensive and long-term exposure	Yes	Yes
2. Canonical upright orientation	Yes	Yes
3. High level expertise	Yes	Yes
4. Predominantly individual, identity level processing	Yes	Yes
5. Featural information critical for identity processing	Yes	Yes
6. Configural information critical for identity processing	Yes	No

(Ge et al., 2006). We therefore concluded that the presence of the inversion effect in the Face condition and its absence in the Chinese Character condition was due to the fact that configural information plays an important role for identifying faces whereas it is not crucial for character identification. Since the participants were all experts in both face and character processing and the Ambiguous Figures stimuli were identical in both conditions, these data provided direct evidence that the differential top–down activation of the face and Chinese character expertise systems determined the presence or absence of the inversion effect during the Ambiguous Figures Task.

Because the face and Chinese character priming tasks in our previous study included both upright and inverted priming stimuli, it is possible that the inversion effect observed in the Ambiguous Figures task in the face condition was due to perseveration of the inversion effect during the face priming tasks. In Experiment 1a of the current study, we use the same priming procedure, but without the inclusion of inverted stimuli in the face and character priming tasks. In order to examine the robustness of this face versus character priming paradigm, we also use a new set of stimuli and include twice as many participants.

2. Experiment 1a

Experiment 1a is a replication of psychophysical evidence that the presence or absence of the face inversion effect can be determined via the top–down activation of face versus Chinese character processing systems.

2.1. Method

2.1.1. Participants
Thirty-two Chinese college students from Zhejiang Sci-Tech University participated (16 males).

2.1.2. Materials and procedure
Two grey-scale averaged Chinese faces were used as prototype faces, one male and one female, to create stimuli for the Real Face Task (Fig. 1). Each averaged face was a composite created from 32 faces of the same race and gender using a standard morphing procedure (Levin, 2000). The two prototype faces were then used to create 24 pairs each of same and different faces by changing the distances between the two eyes and between the eyes and mouth, resulting in faces differing primarily in configural information with the same key face features. A similar procedure was used to create schematic faces for the Schematic Face Task using a single prototype schematic face. Two real Chinese characters were used to create stimuli for the Real Character Task. The first was the same character used by Ge and colleagues, which means "forest." The other was a similarly structured Chinese character meaning "a large body of water." Twenty-four pairs each of same and different Chinese character stimuli were derived by manipulating

the distances between the three elements comprising the prototype Chinese characters similarly to that done for the stimuli used in the Real Face Task. Thus, the resultant characters differed from each other primarily in terms of configuration but with the same key features. This procedure was also used to create nonsense Chinese characters for the Nonsense Character Task using a single prototype nonsense Chinese character. In addition, a set of 24 pairs of ambiguous figures was created for the Ambiguous Figure Task using a similar procedure (Fig. 1).

The Real Face/Real Character tasks each included 48 same and 48 different stimulus pairs (24 distance manipulations × 2 prototypes), while there were 24 pairs of each for the Schematic Face and Nonsense Character tasks (24 distance manipulations × 1 prototype). The stimulus pairs used in the "different" trials differed only in terms of the distance between the three comprising elements. All stimuli were 17.38 × 11.72 cm in size (Fig. 1).

Participants were randomly assigned to either the Face or Character Condition, and performed the Real Face/ Real Character Task first, followed by the Schematic Face/Nonsense Character Task, and finally the Ambiguous Figure Task, with 2 min breaks between tasks. Priming task stimuli were presented only in the upright orientation. However, the Ambiguous Figures were presented in two orientations. In the "upright" orientation, the two elements were above the one element, and in the "inverted" orientation, the one element was above the two elements. Stimuli were presented in random sequence on a 14-in. computer monitor placed 55 cm away from the participant (subtending approximately 17.96 × 12.16 degrees of visual angle). Stimulus presentation and data collection were controlled by experimental software. Participants judged whether the two stimuli were the same or different and received no feedback. They were probed after the experiment about what type of stimuli they believed they were discriminating during the Ambiguous Figures Task. Participants were asked to respond "as accurately and fast as possible."

2.2. Results

Table 2 shows the means and standard deviations of accuracy and correct response latency data for each task in each condition. Paired *t*-tests were performed on the accuracy and latency results between the Face and Charac-

ter Conditions for the two priming tasks. No significant accuracy differences were found. For latency, participants responded significantly faster to real Chinese characters than to real faces, $t(30) = -2.06$, $p < .05$, and significantly faster to nonsense Chinese characters than to schematic faces, $t(30) = -3.57$, $p < .01$.

Repeated measures 2 (Condition: Face vs. Character) × 2 (Orientation: Upright vs. Inverted) ANOVAs were performed on both the accuracy and latency data during the Ambiguous Figures task with Orientation as a repeated measure. There were no significant effects of condition or orientation on accuracy or latency. However, there was a crucial significant orientation-by-condition interaction for the latency data ($F(1, 30) = 6.80$, $p < .05$, $\varepsilon^2 = .19$). Fig. 2 shows that this significant interaction was due to the fact that there was a significant inversion effect for the Face Condition, $t(15) = -3.60$, $p < .01$, but not for the Character Condition, $t(15) = .58$, $p > .05$.

In the post-experiment interview, all participants in the Face Condition reported that they were judging line-drawings of faces in the Ambiguous Figure Task, whereas those in the Character Condition believed they were discriminating "fake" Chinese characters.

2.3. Discussion

As in our previous study (Ge et al., 2006), participants performed as accurately on the Real Face Task as on the Real Character Task, and as well on the Schematic Face as the Nonsense Character Task. In terms of latency during the priming tasks, participants responded faster to real and nonsense Chinese characters than to real and schematic faces. These results are the opposite of those in our previous study, providing evidence that task difficulty during the priming tasks does not account for the observed inversion effects in the Ambiguous Figures Task. Most critically, with exactly the same stimuli, the Ambiguous Figures Task yielded a significant inversion effect in latency in the Face Priming Condition but not in the Chinese Character Condition. Therefore, despite several major methodological changes, the Face and Character priming conditions led to a precise replication of our previous findings in the Ambiguous Figures Task.

At this point, we have consistent evidence from two priming studies that suggests that the presence or absence of the

Table 2
Means (standard deviations) of accuracy (%) and correct response latency (ms) for each task in each condition of Experiment 1a

|	Percent accuracy		Correct latency	
	Face Condition	Character Condition	Face Condition	Character Condition
Step 1. Real Face or Real Character Task (Upright only)	82.7 (12.1)	86.7 (5.8)	3001.00 (1067.61)	2275.48 (915.33)
Step 2. Schematic Face or Nonsense Character Task (Upright only)	91.7 (5.0)	93.2 (5.2)	1872.86 (246.71)	1417.62 (445.90)
Step 3. Ambiguous Figure Task (Upright)				
Upright	93.9 (4.8)	93.5 (4.9)	1670.51 (340.31)	1570.67 (311.74)
Inverted	92.8 (6.9)	93.5 (5.2)	1759.14 (353.33)	1551.21 (300.65)
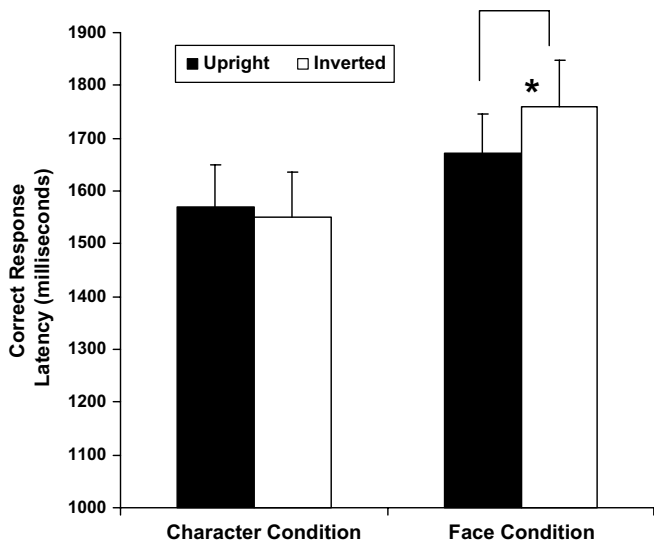
Fig. 2. Mean Correct Response Latencies and standard errors for the Ambiguous Figures in the Face and Character Conditions of Experiment 1a (*p < .05).

face inversion effect can be determined via top–down mechanisms. Specifically, if participants who are experts in processing both Chinese faces and Chinese characters are primed to perceive Ambiguous Figures as faces, then they exhibit a 'face inversion effect' when performing configural discriminations on these stimuli. However, when these subjects are primed to perceive these same Ambiguous Figures stimuli as Chinese characters, they do not exhibit inversion effects when performing configural discriminations on them.

The results of Experiment 1a provide more robust evidence to suggest that the top down activation of the face versus Chinese character expertise systems can determine the presence or absence of the face inversion effect. However, it is not yet clear what role, if any, the activation of the Chinese character expertise system plays during the Ambiguous Figures Task in those participants who participated in the Character priming conditions. In Experiment 1b, we test a group of participants who are not Chinese Character experts, and therefore have not developed a Chinese character expertise system, in the Character Priming Condition in an effort to answer this question.

3. Experiment 1b

Experiment 1b is an examination of the role the Chinese character expertise network played in the differential inversion effects observed in Experiment 1a.

3.1. Method

3.1.1. Participants
Sixteen American college students (6 male) participated.

3.1.2. Materials and procedure
American college students at the University of California, San Diego were tested in the Chinese Character Prim-

ing Condition. None of the participants knew how to read Chinese characters. Other than the fact that these were American students (instead of Chinese students) and that there was no Face Priming Condition, the stimuli, methods, equipment makes and models, computer software, and procedure were identical to those employed in Experiment 1a.

3.2. Results

Table 3 shows the means and standard deviations of accuracy and correct response latency data for each task. Paired *t*-tests were performed on the accuracy results, the latency results for correct trials, and the latency results for both correct and incorrect trials, between the upright and inverted Ambiguous Figures. A significant difference was found in accuracy, with participants performing significantly better when discriminating upright than when discriminating inverted Ambiguous Figures, $t(15) = -2.853$, $p < .01$ (Fig. 3). No significant differences were found for latency during correct trials, $t(15) = 0.97$, $p = .35$ (Table

Table 3
Means (standard deviations) of accuracy (%) and correct response latency (ms) for each task in Experiment 1b (American participants)

	Percent accuracy Character Condition	Correct latency Character Condition
Step 1. Real Character Task (Upright only)	87.63 (6.34)	2492.23 (1058.60)
Step 2. Nonsense Character Task (Upright only)	94.08 (5.25)	1706.17 (452.20)
Step 3. Ambiguous Figure Task		
Upright	93.2 (8.5)	1866.07 (616.93)
Inverted	87.5 (8.5)	1896.80 (591.13)

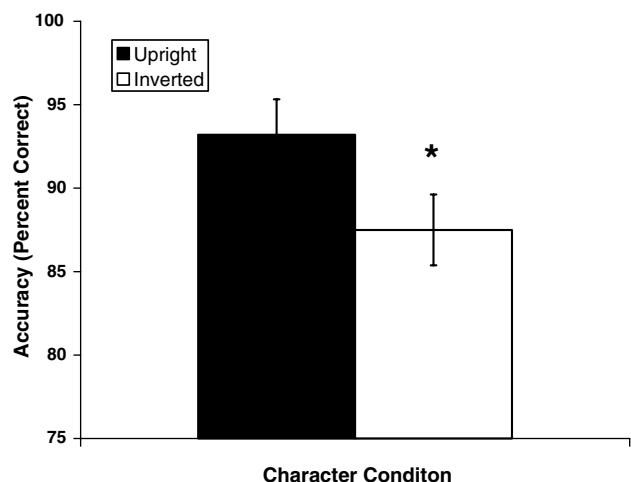Fig. 3. Mean accuracy (%) and standard errors of the mean for the Ambiguous Figures in Experiment 1b (*p < .05).

3). Critically, in the post-experiment interview, all participants reported that they were discriminating either "upright and inverted faces" or "(upright) faces and blocks" during the Ambiguous Figures task. Specifically, the majority of subjects reported perceiving the Ambiguous Figures with two elements on the top and one element on the bottom as "robot-like faces" and the Ambiguous Figures with one element on the top and two elements on the bottom as "inverted (robot-like) faces" or as "blocks."

3.3. Discussion

The results of Experiment 1b indicate that our Chinese character priming tasks were ineffective for priming Chinese character non-experts to perceive the Ambiguous Figures Task stimuli as characters. At the end of the experiment, these participants consistently reported that they perceived the Ambiguous Figures stimuli as "upright faces" and either "inverted faces" or non-face/object stimuli. Consistent with their perceptual report, the data from these participants evidenced a face-like inversion effect when they performed configural discriminations on the Ambiguous Figures. These data contrast with those of the Chinese character experts who participated in the Character Priming Condition of Experiment 1a. Those participants consistently reported perceiving the Ambiguous Figures stimuli as "fake" Chinese characters, and they did not evidence an inversion effect when performing a configural discrimination task on them.

Taken together, these data suggest that our Ambiguous Figures stimuli are ambiguous for participants who are experts in both Chinese character processing and face processing. However, for individuals who are not experts in Chinese character processing, these stimuli do not appear to be ambiguous at all. Instead, they were consistently and robustly perceived as faces in those individuals who lacked a character expertise network.

We attribute this interesting finding to the strength of our Ambiguous Figures stimuli. These stimuli were developed with the explicit purpose of providing a neutral 'balance point' between the face and character expertise systems, while simultaneously embodying key face-related and character-related visual components that would lend them to recognition by, and processing within, each of these systems. In individuals who have both a Chinese character expertise network and a face expertise network, these stimuli are ambiguous and can be perceived as either characters or faces, dependent upon the circumstances (e.g., priming condition). Alternatively, the key face-related visual components consistently and robustly recruited the face expertise network for those individuals who did not have a Chinese character expertise network (regardless of the circumstances).

Because faces and characters were only presented in their upright orientation during the priming conditions, the results of Experiment 1a indicate that the results of

our previous study (Ge et al., 2006) cannot be explained by direct carry-over of poor discrimination of inverted real face and schematic face stimuli into the Ambiguous Figures task. However, it remains possible that the difference in the orientation between the faces (two on top, one on bottom) and Chinese characters (one on top, two on bottom) in the priming tasks of this experiment and the previous experiment (i.e., Ge et al., 2006) may have affected discrimination performance in the Ambiguous Figures condition in these studies. Specifically, if discriminating stimuli with two elements on the top and one element on the bottom is somehow easier than discriminating stimuli with one element on the top and two elements on the bottom, then the practice Chinese character experts in our Chinese character priming condition in Experiment 1a received in discriminating stimuli with one element on the top and two elements on the bottom may have simply negated this effect in the Ambiguous Figures task. In order to control for this possibility, we conducted Experiment 2. In this experiment, we primed Chinese character experts with Chinese character and nonsense character stimuli that had two elements on the top and one element on the bottom, and then examined their performance in discriminating Ambiguous Figures (Fig. 4).
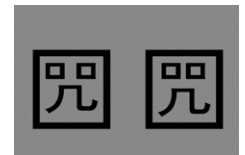

Step 1. Real Character Task (Upright Only)
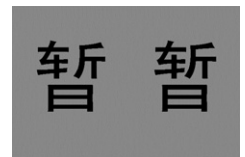
Step 2. Nonsense Character Task (Upright Only)
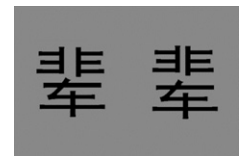
Step 3. Ambiguous Figure Task (Upright)

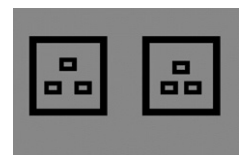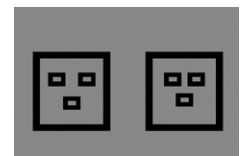**Step 3. Ambiguous Figure Task (Inverted)**

Fig. 4. Design of the Character Condition in Experiment 2.

4. Experiment 2

Experiment 2 is a control for the possibility that orientation-specific learning effects may have negated inversion effects in Experiment 1a.

4.1. Method

4.1.1. Participants
A new group of 16 Chinese college students (8 male) from Zhejiang Sci-Tech University participated.

4.1.2. Materials and procedure
Chinese college students were tested in a Chinese character priming condition where the priming stimuli (Chinese characters and nonsense Chinese characters) had two elements on the top and one element on the bottom (Fig. 4). Other than these changes in the stimuli, the stimuli and methods were identical to those employed in the Character Priming Conditions of Experiments 1a and 1b.

4.2. Results

Table 4 shows the means and standard deviations of accuracy and correct response latency data for each task in each condition. Paired *t*-tests were performed on the accuracy and latency results between the upright and inverted Ambiguous Figures. No significant differences were found (Accuracy: $t(15) = 0.46$, $p = .65$; Latency: $t(15) = 0.62$, $p = .55$). In the post-experiment interview, all participants reported that they were discriminating "fake" Chinese characters during the Ambiguous Figures task.

4.3. Discussion

The results of Experiment 2 indicate that the difference in the orientation between the faces (two on top, one on bottom) and Chinese characters (one on top, two on bottom) during the priming conditions cannot explain the significant inversion effect in the face condition but not in the character condition in Experiment 1a. Together, the results

of these experiments provide more conclusive evidence to support the hypothesis that the top–down activation of a face processing expertise system that has a specific sensitivity to upright versus inverted configural information for processing plays a crucial role in the face inversion effect.

From a developmental perspective, the face and character expertise systems are the product of the visual system's extensive experience with processing a multitude of individual faces and characters, respectively. Because the processing demands are different for faces and characters (see Table 1), the systems are tuned differentially to meet the demands of face or character processing. Thus, the face expertise system is tuned to be disproportionately sensitive to upright as compared to inverted configural information, whereas the character expertise system is not.

Why are the processing demands for configural information different for faces and Chinese characters? After all, characters are highly similar to faces on many dimensions. Like faces, characters are omnipresent in Chinese societies and Chinese people are exposed to them from early childhood. Also, like faces, characters have a canonical upright orientation, and are processed primarily at the individual character level (because each character carries specific meanings). Literate Chinese adults are experts at processing thousands of individual characters as well as they do with faces. Furthermore, both faces and Chinese characters contain featural and configural information.

Why, then, is upright configural information special for face processing but not for character processing? One possibility is that the difference is due to different top–down cognitive strategies one has learned to use for processing upright faces versus upright characters. Experience may have taught one to focus on configural information for upright face processing but to focus on featural information for upright character processing. Specifically, to recognize the identity of characters in different fonts and handwriting one must learn to ignore the within-class variation in the configuration of the parts, whereas face configural information is relatively consistent within an individual and, therefore, provides critical information for identity discrimination.

A possible alternative explanation is that the reason for the inversion effect in one set of stimuli versus the other is that the configural variations in inverted characters have a closer match with the configural variations in upright characters, but inverted faces do not have similar variations in their configurations as upright faces. This would imply that the perceptual representation for characters is already well-tuned for representing configural variations in inverted characters, but the perceptual representation for faces is not well tuned for configural variations in inverted faces. In this account, the differences arise at the level of perceptual representation rather than the presumably later level of identification.

To date, no evidence exists that determines the validity of these two possibilities. It has been long proposed that face processing expertise systems develop through extensive

Table 4
Means (standard deviations) of accuracy (%) and correct response latency (ms) for each task in behavioral Experiment 2

	Percent accuracy Character Condition	Correct latency Character Condition
Step 1. Real Character Task (Upright only)	85.9 (7.9)	1930.33 (623.28)
Step 2. Nonsense Character Task (Upright only)	92.7 (7.0)	1813.61 (740.75)
Step 3. Ambiguous Figure Task		
Upright	94.9 (5.0)	1633.57 (436.29)
Inverted	95.3 (5.5)	1657.37 (392.28)

experience discriminating human faces (see Bukach, Gauthier, & Tarr, 2006; Gauthier & Nelson, 2001, for reviews). Although research has shown that with experience children become increasingly more reliant on configural information for face processing (Friere & Lee, 2004), little is known as to why experience should engender such developmental change and what perceptual-cognitive mechanisms are responsible for this change. To bridge the gap in the literature, we employ a neurocomputational model of face and object recognition that has previously been shown to account for a number of important phenomena in facial expression processing, holistic processing, and visual expertise (Cottrell, Branson, & Calder, 2002; Dailey & Cottrell, 1999; Dailey, Cottrell, Padgett, & Adolphs, 2002; Joyce & Cottrell, 2004; Tran, Joyce, and Cottrell, 2004; Tong, Joyce, & Cottrell, 2005; Tong, Joyce, & Cottrell, in press). Specifically, we train neural network models of face and Chinese character expertise and test them on the Ambiguous Figures Task from Experiment 1. One of the nice properties of explicit models with levels of representation is that one can identify where such an effect might first arise in the model, and therefore distinguish between the two possibilities outlined above.

5. Experiment 3

5.1. Modeling inversion effects with computational models

Our application of the model involves employing the same initial processing mechanisms for both face and character networks, and then making comparisons between models that are trained separately for Chinese face versus Chinese character identity discrimination expertise. The training of two separate networks for these two tasks is motivated by neuroimaging experiments that have shown that face processing tasks and Chinese character processing tasks activate different brain regions (e.g., Chen, Fu, Iversen, Smith, & Matthews, 2002; Haxby, Hoffman, & Gobbini, 2002). Hence we assume that the top–down priming of seeing the ambiguous figures as faces or as characters will invoke the use of one or the other of these networks. If the nature of the visual input to the discrimination task is what determines sensitivity to configural information, then the physical characteristics of the input to each network will interact with the neural network model to produce differential inversion effects. Specifically, we predict that training the network on Chinese faces will result in inversion effects but training the network on Chinese characters will not.

5.2. A computational model of classification

As described above, we model the two conditions of the experiment, face and character priming, using two neural networks, a face identification network and a character identification network (see Fig. 5). These two networks start with the same preprocessing, using Gabor filters (modeling complex cell responses in V1), and then diverge into two processing streams. The hypothesis is that in the face condition, the subjects are using their face processing system (seeing the ambiguous characters as faces), and in the character condition, they are using their character processing system (seeing the ambiguous characters as characters). The next layers of the model perform two principal component analyses (PCA), one for characters and one for faces, to extract covariances among the Gabor filter responses, and to reduce the dimensionality of the data. This processing can also be carried out by a neural network (Cottrell, Munro, & Zipser, 1987; Sanger, 1989), but we simply use PCA here. The responses of the principal components are given as input to a hidden layer trained by back propagation to extract features useful for discriminating the inputs into the required classes. We now describe each of the components of the model in more detail.
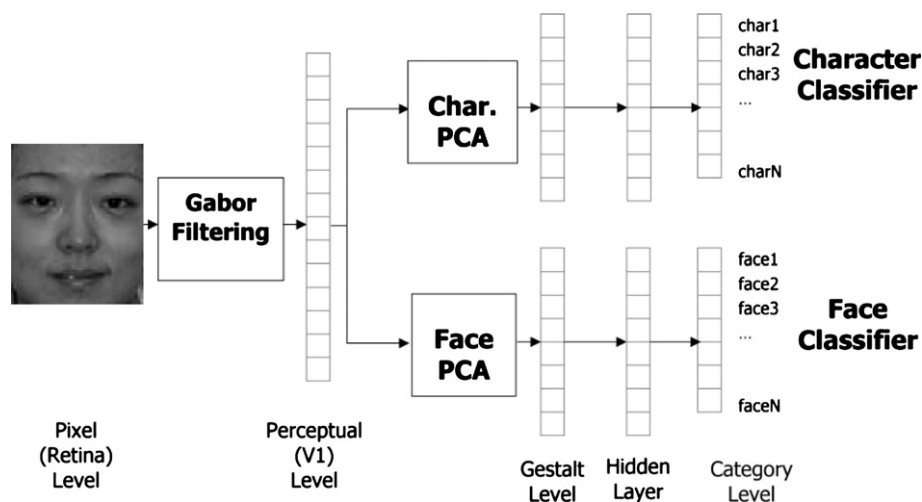

Fig. 5. Visual representation of the structure of the neurocomputational model.

5.2.1. Perceptual layer of V1 cortex

Research suggests that the receptive fields of striate neurons are restricted to small regions of space, responding to narrow ranges of stimulus orientation and spatial frequency (Jones & Palmer, 1987). Two dimensional Gabor filters (Daugman, 1985) have been found to fit the spatial response profile of simple cells quite well. We use the magnitudes of quadrature pairs of Gabor filters, which models complex cell responses. The Gabor filters span five scales and eight orientations.

We used biologically realistic filters following (Dailey & Cottrell, 1999; Hofmann, Puzicha, & Buhmann, 1998), where the parameters are based on those reported in Jones and Palmer (1987), to be representative of real cortical cell receptive fields. The basic kernel function is:

$$G(\vec{k}, \vec{x}) = \exp i \vec{k} \cdot \vec{x} \exp \left(-\frac{k^2 \vec{x} \cdot \vec{x}}{2\sigma^2} \right)$$

where

$$\vec{k} = [k \cos \phi, k \sin \phi]$$

and k controls the spatial frequency of the filter function G. x is a point in the plane relative to the wavelet's origin. Φ is the angular orientation of the filter, and σ is a constant. Here, $\sigma = \pi$, Φ ranges over $\{0, \pi/8, \pi/4, 3\pi/8, \pi/2, 5\pi/8, 3\pi/4, 7\pi/8\}$, and

$$K_i = \frac{2\pi}{N} 2^i$$

where N is the image width and i an integer. We used 5 scales with $i = [1,2,3,4,5]$.

Thus every image (50 by 40 pixels) is presented by a vector of $50 \times 40 \times 5 \times 8 = 80{,}000$ dimensions.

5.2.2. Gestalt layer

In this stage we perform a PCA of the Gabor filter responses. This is a biologically plausible means of dimensionality reduction, since it can be learned in a Hebbian manner (Sanger, 1989). PCA extracts a small set of informative features from the high dimensional output of the previous perceptual stage. It is at this point that the processing of characters and faces diverge: We performed separate PCA's on the face and character images, as indicated in Fig. 6. This separation is motivated by the fact that Chinese character processing and face processing are neurologically distinct (Chen et al., 2002; Haxby et al., 2002; Kanwisher, McDermott, & Chun, 1997). In each case, the eigenvectors of the covariance matrix of the patterns are computed, and the patterns are then projected onto the eigenvectors associated with the largest eigenvalues. At this stage, we produce a 50-element PCA representation from the Gabor vectors for each class. Before being fed to the final classifier, the principal component projections are shifted and scaled so that they have 0 mean and unit standard deviation, known as z-scoring (or whitening).

5.2.3. Categorization layer

The classification portion of the model is a two-layer back-propagation neural network with fifty hidden units. A scaled tanh (LeCun, Bottou, Orr, & MÄu ller, 1998) activation function is used at the hidden layer and the softmax activation function was used at the output level. The network is trained with the cross entropy error function (Bishop, 1995) to identify the faces or characters using localist outputs. A learning rate of 0.02, a momentum of 0.5 and training length of 100 epochs were used in the results reported here.

5.3. Methods

5.3.1. Image sets

We used training images of 87 Chinese individuals (45 females), 3 expressions each (neutral, happy, and angry), for a total of 261 images, to train the face networks on identity discriminations. Face images are triangularly aligned and cropped (Zhang & Cottrell, 2004) to 50 by 40 pixel images. That is, the triangle formed by the eyes and mouth is shifted, rotated, and scaled to be as close as possible in Euclidean distance to a reference triangle. This preserves configural information in the images, while controlling for face size and orientation. Eighty-seven Chinese characters, 3 fonts each (line drawing, hei, and song) were used to train the character networks on identity discriminations (Fig. 6).

Ten images (5 upright/5 inverted) were used as Ambiguous Figure images for the final tests of the research hypotheses. An upright prototype image was created by matching the centers of the squares to the average position of features (eyes and mouth) in the face training set. The line width matched that of the character training set. Four other upright images were created by moving the upper two squares 2 pixels in/out and the lower one square 2 pixels up/down. The inverted ambiguous images were created by simply flipping the upright images upside down.

5.3.2. Initial evaluation of network performance

First we evaluated how well our model performed on the classification task on faces and characters respectively. Cross-validation was used for evaluation, i.e. for the face network, two expressions (e.g., happy, angry) were used to train the network and one expression (e.g., neutral)
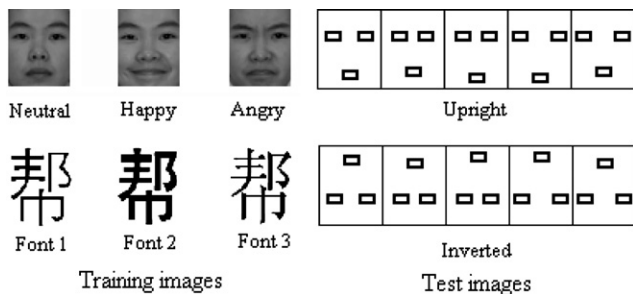

Fig. 6. Example training images (left) and test images (right) used in the neurocomputational modeling experiment.

Table 5
Accuracy evaluation on the validation set in modeling experiment

Face	Neutral	Happy	Angry
Accuracy (%)	98.64	97.05	90.73
Standard error (%)	0.24	0.26	0.51
Character	Font 1	Font 2	Font 3
Accuracy (%)	97.63	96.62	92.10
Standard error (%)	0.31	0.36	0.57

was used as a test set to see how well the learning of identity on the other two expressions generalized to this expression; for the character network, two fonts were used as training set and one font as a test set. Table 5 shows performance on the test sets for the face and character networks respectively (averaged over 16 networks each condition).

The networks achieved over 90% accuracy for every condition. This is remarkable given that only image images were used for training for each individual face or character. This classification rate was sufficient to show that our models represented face/character images well. With this result in hand, we proceeded to model the experimental condition.

5.3.3. Modeling discriminability

For the following experiments, we simply trained the network on all 261 face/character images, since we were only interested in obtaining a good face/character representation at the hidden layer. All the other parameters remained the same as in the above evaluation experiments.

Hidden unit activations were recorded as the network's representation of images. In order to model discriminability between two images, we presented an image to the network, and recorded the hidden unit response vector. We then did the same with a second image. Finally, we modeled similarity as the correlation between the two representations, and discriminability as one minus similarity. Note that this measure may be computed at any layer of the network, including the input image, the Gabor filter responses, and the PCA representation, as well as the hidden and output responses. We computed the average discriminability between image pairs in character networks and face networks. This average was computed over 16 face/character networks that were all trained in the same manner but with different initial random weights.

5.4. Results

Fig. 7 presents the discriminability of the face networks and the character networks on the upright and inverted ambiguous image sets, averaged over 16 networks (matching the number of human subjects tested in Experiment 1a), respectively. For the face networks, the discriminability of the upright images was significantly better than that of the inverted image set ($F = 53.23$, $p < .01$). However, the character networks do not exhibit an inversion effect ($F = 1.64$, $p = .211$).
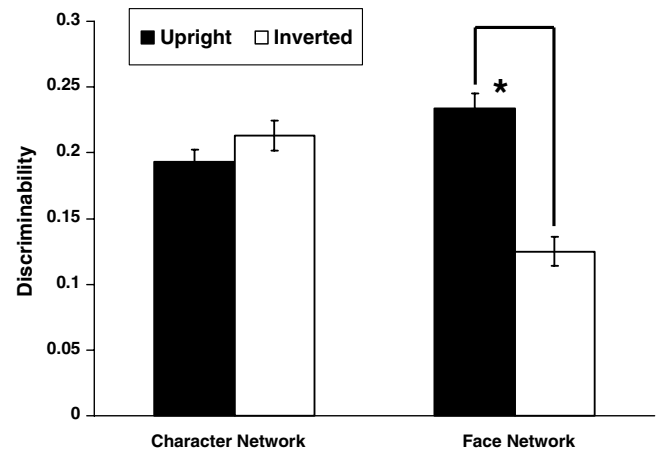

Fig. 7. Mean discriminability and standard errors of the mean for the Ambiguous Figures in the Face and Character Conditions of neurocomputational modeling experiment (*$p < .05$).

5.4.1. Discriminability at different stages of the model

Where do these effects come from? Recall our definition of discriminability: one minus similarity, where similarity is equal to the correlation between representations. Hence, we can assess similarity and discriminability at other stages of processing, i.e., raw images, Gabor filter responses, PCA (Fig. 5). Note that for preprocessing stages (raw images, Gabor filter responses), the discriminability of the upright ambiguous image set equals that of the inverted set since the one is the upside down version of the other, and the orientations of the Gabor filters span this rotation. That is, the representation is not 'tuned' in any way to upright versus inverted images—both are represented equally well. However, at the PCA level, the statistics of the upright images constrain the representations. Recall there are two PCA's, one for characters and one for faces. Both PCA's are tuned to upright versions of these stimuli—they are extracting the dimensions of maximum variance in the Gabor filter responses to the upright images in each set. Fig. 8 shows that at the PCA level there is a much larger gap between the discriminability of the upright ambiguous image set and that of the inverted set in the face PCA's than in the character PCA's. Furthermore, the gap in the face networks at the PCA level is carried on to the hidden layer of the category level, but the small difference in the character networks turned out to be non-significant at the hidden level. Since the PCA is only performed on upright images, it is not surprising that differences between upright and inverted discriminability arise. However, the difference in the size of the inversion effect for the ambiguous figures in the face versus character networks must be due to the inability of the principal components of faces to represent the variability in "upside down" ambiguous characters, while the differences between upright and inverted characters appears to match better in the two conditions. That is, the variance of the upright ambiguous images correlates more with the variance in upright face
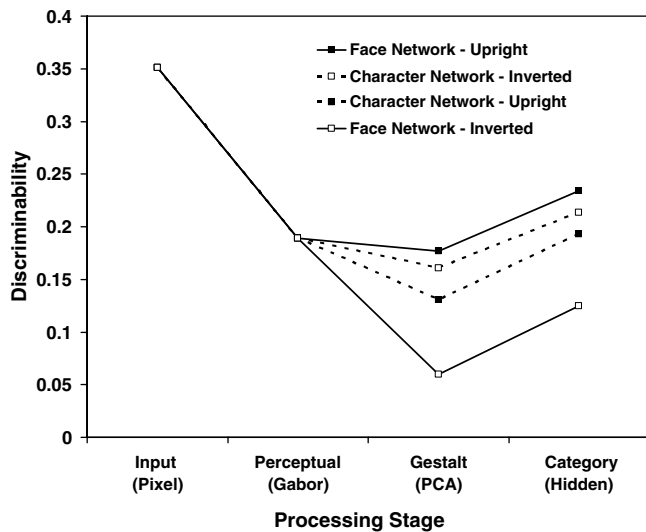
Fig. 8. Discriminability of upright and inverted Ambiguous Figures stimuli at the different stages of the neurocomputational model in the face and character conditions.

images than the variance of inverted ambiguous images correlates with upright face images.

5.5. Discussion

The results of the application of our neural network model to Chinese face and Chinese character expertise networks show that a face expertise network produces a significant decrease in discrimination performance upon Ambiguous Figures involving configural discriminations when they are inverted, whereas the Chinese character expertise network does not produce such an effect. These findings support the hypothesis that the specialized expertise with configural information acquired through extensive training of the visual system with discriminating upright faces is *sufficient* to explain the face inversion effect. Furthermore, examination of discrimination at each of the stages of the two models indicates that these effects emerge at the level of PCA processing, which is designed to model the structural representation of the stimuli.

An important fact to remember is that we performed two PCA's: One for faces and one for Chinese characters. This decision was made because face and Chinese character processing are believed to be neurally distinct, with faces typically activating right hemisphere regions more strongly than Chinese characters, and vice-versa. The fact that we compute separate PCA's in these two cases reflects the hypothesis that there is an *early divergence* of representation in cortex for these two classes of stimuli. Computationally, this means that the PCA level is the first level at which there is a difference in processing by the model.

The goal of PCA is to account for the most variability in the data for the given number of components. The fact that there is a big difference in the discriminability between upright and inverted ambiguous characters at the PCA level for a face PCA means that the upright ambiguous stimuli better match the kind of variability found in the face data than the inverted ambiguous stimuli. Presumably, this variability is due to differences in the spacing between the eyes. On the other hand, there appears to be less overall variability in characters. While this may seem to be a purely bottom–up phenomenon, the fact that the PCA's are separated at all means that there is some supervision going on: the data is being divided based on class before the statistics are computed. What is remarkable, however, is that there is no need for an expertise effect here. There is simply an effect of separating the data into two classes. That is, this kind of learning (of the principal components) is thus semi-supervised, based only on the class of the data, but is not based upon separating the data at the subordinate level. However, in order for these data to match the data in the psychophysical experiment, there must have been a top–down effect of treating these stimuli as faces versus treating these stimuli as characters. Otherwise, the difference would not have arisen.

The fact that this separation is maintained at the hidden unit level means that these configural differences are important for distinguishing between different faces. There is numerically (but not significantly) less maintenance of this difference in the character condition. This suggests that configural differences are less important for discriminating the identities of Chinese characters, consistent with our hypothesis.

To further investigate the role of discrimination expertise, we re-ran the entire experiment using only a single PCA for both characters and faces, as we have done in previous studies of visual expertise (Joyce & Cottrell, 2004). This represents the notion that there might be a single, unsupervised representational space for both faces and Chinese characters before learning of the two domains is performed. We refer to this as a *late divergence* model. In this version of the experiment, we did not find a significant inversion effect using either the PCA layer or the hidden layer of the face expert network (data not shown). These data suggest that the early divergence model, which involves distinct face and Chinese character networks, is the correct model.

6. General discussion

We used psychophysical priming experiments and a neurocomputational model to examine the roles of visual discrimination expertise and visual input in the face inversion effect. The psychophysical priming experiments showed that the effect of inversion on Chinese adults' processing differed for faces and Chinese characters. Chinese participants showed a significant inversion effect during the processing of the Ambiguous Figures after the face-priming tasks. However, after character-priming tasks, the same Ambiguous Figures did not produce a significant inversion effect. The results of additional control experiments provided further support for the critical role of the activation of the face versus Chinese character expertise

systems in eliciting the face inversion effect in our priming paradigm. Taken together, this series of experiments provides more conclusive evidence that the top–down activation of a face processing expertise system plays a crucial role in the face inversion effect. Furthermore, we have shown that the activation of this system during configural processing is sufficient to account for the face inversion effect.

We then applied a neurocomputational model to simulate the discrimination of upright and inverted Ambiguous Figures by neural networks trained to discriminate either faces or Chinese characters with a high level of expertise. The neurocomputational modeling also showed a significant inversion effect when discriminating the Ambiguous Figures in the face condition but not in the character condition. These results thus replicated the results of the behavioral experiments. This modeling experiment also provides the first evidence to suggest that the specialized expertise with upright configural information that develops through extensive training of the visual system with upright faces is sufficient to explain the development of the face inversion effect. Furthermore, the fact that this difference emerges at the PCA level, where training is unsupervised, suggests that this is a bottom up effect. However, our model suggests that, at some point in the visual system, some process must have decided to "send" faces to one system and send characters to another. In this sense, there is some supervision involved, in the choice of stimuli from which to extract covariances. This process is not implemented in our model, but previous models using a "mixture of experts" architecture have shown the ability to do this (Dailey & Cottrell, 1999). Regardless, the results of the modeling in the current study support the hypothesis that the nature of the visual input to the identity discrimination task (i.e., upright faces or upright Chinese characters) is what determines the neural network's sensitivity to configural information. In the context of the current study, the nature of the visual input also determined the presence or absence of inversion effects.

The face inversion effect is an important marker for face-specific processing that has been studied extensively to further our understanding of face processing for more than three decades (e.g., Yin, 1969; see Rossion & Gauthier, 2002, for review). The results of these studies suggest that the inversion effect is primarily due to the processing of configural information in faces (e.g., Farah et al., 1995; Freire et al., 2000; but see Yovel & Kanwisher, 2004). Although the results of several recent studies provide indirect evidence that the high levels of visual discrimination expertise associated with face processing may explain these phenomena (e.g., Gauthier & Tarr, 1997), very little is known about the direct relationships of visual expertise and visual input to face configural processing and the inversion effect.

Because Chinese characters are similar to faces in many important regards, including their level of visual discrimination expertise (see Table 1), but do not require configural information for identity discrimination, they are ideal comparison stimuli for experiments aimed at understanding the roles of visual expertise and visual input to face configural processing. Previous research has shown that with experience children become increasingly more reliant on configural information for face processing (see Friere & Lee, 2004 for review). However, these studies have not provided information regarding why experience should engender such developmental change or what perceptual-cognitive mechanisms are responsible for this change. Using neural network modeling to simulate the learning process during the formation of the face versus Chinese character expertise systems, we explored how the visual system might change as it learns to extract invariant information from either faces or Chinese characters in the most optimal manner. Despite identical initial visual processing mechanisms and highly similar levels of visual training and expertise, these two models performed differently on the discrimination of the exact same stimuli, the Ambiguous Figures. The face network performed significantly better on configural discriminations on the upright stimuli than on the inverted stimuli. This is because the interaction among the physical properties of faces and the properties of the modeled neuronal networks of the visual system resulted in an expertise system specifically tuned to upright configural information. On the other hand, the character network performed the same on both the upright and inverted stimuli because the learning process associated with the development of this expertise system resulted in a neural network that was equally capable of using inverted and upright configural information, even though the training characters were all presented upright. These results indicate that the visual learning process associated with the formation and development of the face expertise and character expertise networks is sufficient to explain why increasing expertise with faces results in specialized sensitivity to the discrimination of upright configural information, and the face inversion effect, but character expertise processing does not.

Neurocomputational models have previously been applied to face processing in an effort to further our understanding of face processing and expertise. In the realm of featural versus configural processing, Schwaninger, Wallraven, and Bulthoff (2004) compared the performance of a face processing network that involved two stages of processing. In this model, a 'component' processing stage where features were identified and a 'configural' processing stage where the relationship of these features to one another was examined. The results showed that the model's performance was highly similar to the performance of human subjects on a number of psychophysical tasks, suggesting the possibility that these two mechanisms may in fact be separate. In 2006, Jiang and colleagues provided evidence that their biologically plausible, simple shape-based model, that does not explicitly define separate 'featural' and 'configural' processes, can account for both featural and configural processing, as well as both face

inversion effects and inversion effects associated with non-face object expertise (Jiang et al., 2006). Of particular note is the fact that featural and configural processing in this study were both well accounted for by the same shape-based discrimination mechanisms, and that increasing the simple shape-based neurocomputational model's training experience with non-face objects resulted in face-like inversion effects.

More recently, we employed the same biologically plausible neurocomputational model presented in the current study to test the hypothesis that the shift from a more feature-based face processing strategy to a combined featural and configural processing strategy over development could be explained by increased experience with faces. Specifically, our results suggested that as the number of faces one is exposed to increases, the relative reliance on configural versus featural information increases, as well (Zhang & Cottrell, 2006). In the current study, we extend this line of work by training two separate neurocomputational models on two different types of stimuli and comparing their performance in an experimental task. Here, we tested the hypothesis that extensive visual discrimination experience with faces, but not extensive visual discrimination expertise with Chinese characters, results in orientation-specific configural sensitivity and a face inversion effect. Because configural information is not tied to identity in Chinese characters, these experimental data provide critical new insight into the specifics of the relationships of visual expertise and visual input to face configural processing and the inversion effect.

In sum, the current psychophysical and neurocomputational modeling experiments provide evidence that different visual expertise systems for face versus Chinese character processing are formed through a dynamic yet mechanistic learning process that is the result of the unique properties of the human visual system in combination with extensive experience making visual discriminations on these groups of stimuli that each have unique physical characteristics. Then, once these expertise systems are formed, top–down mechanisms can influence when and how they are recruited. These data provide important new insights into the roles of visual expertise and visual input in the face inversion effect, and provide a framework for future research on the neural and perceptual bases of face and non-face visual discrimination expertise.

Acknowledgments

References

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

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Science, 10*(4), 159–166.

Chen, Y., Fu, S., Iversen, S. D., Smith, S. M., & Matthews, P. M. (2002). Testing for dual brain processing routes in reading: a direct contrast of Chinese character and Pinyin reading using fMRI. *Journal of Cognitive Neuroscience, 14*, 1088–1098.

Cottrell, G. W., Branson, K. M., & Calder, A. J. (2002). Do expression and identity need separate representations? In *Proceedings of the 24th annual cognitive science society conference*, Mahwah, NJ. Cognitive Science Society.

Cottrell, G. W., Munro, P., & Zipser, D. (1987). Learning internal representations from gray-scale images: An example of extensional programming. In *Proceedings of the ninth annual cognitive science society conference*, Seattle, WA. pp. 461–473.

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, spacial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of American A, 2*, 1160–1169.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 628–634.

Friere, A., & Lee, K. (2004). Person recognition by young children: Configural, featural, and paraphernalia processing. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood: Current perspectives* (pp. 191–205). New York: Nova Science Publishers.

Freire, A., Lee, K., & Symons, L. (2000). The face inversion effect as a deficit in the encoding of configural information: direct evidence. *Perception, 29*, 159–170.

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology, 11*(2), 219–224.

Gauthier, L., & Tarr, M. J. (1997). Becoming a "greeble" expert: exploring mechanisms for face recognition. *Vision Research, 37*, 1673–1682.

Ge, L., Wang, Z., McCleery, J., & Lee, K. (2006). Activation of face expertise and the inversion effect. *Psychological Science, 17*(1), 12–16.

Haxby, J. B., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry, 51*, 59–67.

Hofmann, T., Puzicha, J., & Buhmann, J. M. (1998). Unsupervised texture segmentation in a deterministic annealing framework. *IEEE PAMI, 20*(8), 803–818.

Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron, 50*, 159–172.

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

Joyce, C., & Cottrell, G. W. (2004). Solving the visual expertise mystery. In *Proceedings of the neural computing and psychology. Workshop 8, progress in neural processing*. London, UK: World Scientific.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specializd for face perception. *Journal of Neuroscience, 17*(11), 4302–4311.

LeCun, Y., Bottou, L., Orr, G. B., & MÄuller, K.-R. (1998). Efficient backprop. In *Neural networks: Tricks of the trade. Lecture Notes in Computer Sciences* (Vol. 1524, pp. 5–50). Springer.

Levin, D. T. (2000). Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129*(4), 559–574.

Rossion, B., & Gauthier, L. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews, 1*, 62–74.

Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feed forward neural network. *Neural Network, 2*, 459–473.

Schwaninger, A. C., Wallraven, C., & Bulthoff, H. H. (2004). Computational modeling of face recognition based on psychophysical experiments. *Swiss Journal of Psychology, 63*(3), 207–215.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 904–915.

Tong, M., Joyce, C., & Cottrell, G. W. (2005). Are greebles special? Or, why the fusiform fish area (if we had one) would be recruited for sword expertise. In *Proceedings 27th annual cognitive science society conference*, La Stresa, Italy. Cognitive Science Society.

Tong, M. H., Joyce, C. A., & Cottrell, G. W. (in press). Why is the fusiform face area recruited for novel categories of expertise? A neurocomputational investigation. *Brain Research*.

Tran, B., Joyce, C. A., & Cottrell, G. W. (2004). Visual expertise depends on how you slice the space. In *Proceedings of the 26th annal cognitive science society conference*. Chicago, Illinois. Cognitive Science Society.

Yin, R. K. (1969). Looking at upside–down faces. *Journal of Experimental Psychology, 81*, 141–145.

Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron, 44*, 889–898.

Zhang, L., & Cottrell, G. W. (2004). When holistic processing is not enough: Local features save the day. In *Proceedings of the twenty-sixth annual cognitive science society conference*.

Zhang, L., & Cottrell, G. W. (2006). Look Ma! No network: PCA of gabor filters models the development of face discrimination. In *Proceedings of the 28th annual cognitive science conference, Vancouver, BC, Canada*. Mahwah: Lawrence Erlbaum.