Cerebral Cortex June 2010;20:1314–1322
doi:10.1093/cercor/bhp195
Advance Access publication September 18, 2009

The Neural Basis of Categorical Face Perception: Graded Representations of Face Gender in Fusiform and Orbitofrontal Cortices

Jonathan B. Freeman[1], Nicholas O. Rule[1], Reginald B. Adams Jr[2] and Nalini Ambady[1]

[1]Department of Psychology, Tufts University, Medford, MA 02155, USA and [2]Department of Psychology, The Pennsylvania State University, University Park, PA 16802, USA

Face gender, like many other things, is perceived categorically: Subjective perceptions are distorted toward the categories, male or female, and the objective gradiency inherent across faces is partially lost. The neural basis of such categorical face perception remains virtually unknown. Participants passively viewed faces whose sexually dimorphic content was morphed monotonically from male to female while neural activity was measured using functional magnetic resonance imaging. Subjective perceptions revealed strong nonlinearity despite monotonic linear changes in face gender, consistent with categorical perception. Neuroimaging results indicated that the lateral fusiform gyrus, bilaterally, and the fusiform face area linearly encoded graded parameters of objective face gender, but these regions correlated substantially less with subjective perceptions (which were nonlinear and affected by categorical perception effects). Such subjective perceptions, however, were represented in the orbitofrontal cortex, but this region correlated substantially less with objective parameters. The attention-independent graded representations of face gender in fusiform and orbitofrontal cortices reveal how objective face parameters are encoded and transformed into subjective categorically warped perceptions in the human brain.

Keywords: categorical perception, face gender, face perception, functional magnetic resonance imaging, fusiform cortex, orbitofrontal cortex

Introduction

If every stimulus were perceived as a novel experience for humans, we would quickly become inundated with a bewildering amount of redundant information. Thus, the human brain groups stimuli with similar characteristics into a single category (Rosch 1978; Murphy 2002). Whether innate or acquired, this can be accomplished by the perceptual system differentially expanding and compressing the signal of the external world, where differences among stimuli on one side of a category boundary are attenuated at the same time that analogous differences among stimuli straddling the boundary are amplified. This describes the *categorical perception* phenomenon, which has been theorized to play a central role in categorization and category learning (Harnad 1987).

A telltale sign of categorical perception is that despite perceptual input changing monotonically along a continuum, stimuli are nevertheless perceived with a discrete category boundary occurring at a certain location along the continuum. If perceivers are asked to identify stimuli (e.g., colors: green vs. yellow), plotting the proportion of identifications (e.g., the likelihood of perceiving green) as a function of monotonic *linear* changes in the stimuli would reveal a strong *non-linear* effect, typically a step-like sigmoidal function: Changes along the continuum on one side of a category boundary have virtually no influence on identification, whereas those straddling the boundary give rise to an abrupt perceptual shift, for example, green suddenly becomes yellow (Harnad 1987). For instance, a rainbow is made up of monotonic changes in light wavelength; yet, we perceive this continuum as discrete bands of colors with sharp category boundaries (Bornstein and Korda 1984). This phenomenon is not limited to visual perception. For instance, the difference between hearing the phoneme, "bah" (/ba/), versus the phoneme, "pah" (/pa/), is a characteristic of speech articulation referred to by psycholinguists as voice onset time (VOT). As VOT monotonically rises from 0 to 40 ms, between 15–20 ms participants suddenly switch between hearing "bah" versus "pah" despite the fact that on either side of this boundary participants are highly consistent in their perceptions (Liberman et al. 1957; McMurray et al. 2003).

Recently, research in categorical perception has been extended to faces. Perceivers readily and rapidly extract a variety of information from another's face, including identity, emotional status, gender, race, and age, among others (Macrae and Bodenhausen 2000). Prior evidence suggests that our ability to perceive this information is made possible by a categorical perception mechanism, similar to that involved in colors and phonemes. For instance, using familiar faces, perceivers were found to be better at distinguishing the identity of a pair of faces that belonged to different people than when they belonged to the same person, even though the actual physical difference between them was made identical using a morphing algorithm (Beale and Keil 1995). Using photographs of twins, Stevenage (1998) found a similar effect and also showed that this phenomenon can be acquired across the course of a task. This provides evidence that we perceive others' face identity in categorical fashion and the ease with which it may be acquired.

Several studies have also provided evidence for categorical perception effects in recognizing face emotions, showing that perceivers identify emotional expressions in discrete fashion although the perceptual signal may be continuous (e.g., Etcoff and Magee 1992; Calder et al. 1996). Most recently, the phenomenon has been extended to face gender. Campanella et al. (2001) presented participants with unfamiliar faces morphed monotonically along gender (i.e., from male to female). Consistent with a categorical perception effect, participants' identification function exhibited a discrete category boundary at critical changes along the perceptual continuum where a male face suddenly was perceived to be female. This converged with additional evidence from a discrimination task. Moreover, these categorical perception

effects were not obtained for inverted faces, which disrupt configural face processing and the extraction of gender (Valentine 1988; Campanella et al. 2001). By showing categorical perception effects for upright faces and not inverted faces, the authors showed that their evidence for categorical perception was indeed due to gender processing rather than some task constraint or stimulus-related confound (Campanella et al. 2001). Thus, like emotional expressions, face identity, colors, and phonemes, face gender appears to be perceived in an inherently categorical fashion.

Although the way in which the human brain perceives face gender remains largely unknown, the neural basis of face perception has been studied in great detail. Among other face-sensitive regions in occipitotemporal cortex, one region in particular, the fusiform face area (FFA), has been repeatedly put forth as a critical player. Despite competing theories regarding this region's status as a face-specializing module versus more generic perceptual mechanism, researchers agree that it plays an important role in perceiving others' faces (Kanwisher and Yovel 2006). Haxby et al. (2000) have theorized a distributed neural system for face perception, including the FFA and other regions in temporal cortex, in addition to their wider interactions with frontal, parietal, and subcortical regions. Their model proposes a dual processing route, where first the early perception of facial features is mediated by the inferior occipital gyrus, which then divides the labor onto the lateral fusiform gyrus (FG), primarily treating static structural cues (i.e., identity), and the superior temporal sulcus, primarily treating dynamic expressions (i.e., emotions). Although not espoused by all (e.g., Calder and Young 2005), the dual processing model has been highly influential, leading to an extensive line of work testing the dissociable neural processing of face identity versus emotional expressions (e.g., Winston et al. 2004).

Previous studies have compared neural responses to male versus female faces (Fischer et al. 2004) or examined responses during explicit forced-choice gender classification tasks with male and female faces (e.g., Paller et al. 2003). Studies of this latter type were not primarily interested in gender classification but instead used it as a control task for another process of primary interest, such as memory retrieval in person recognition (e.g., Paller et al. 2003). Here, rather than examine the active classification of gender, we were interested in how it is implicitly and automatically encoded.

Acknowledging that face gender is perceived categorically, one question of interest is how monotonic objective changes along a continuum of sexually dimorphic face content (the diagnostic perceptual information that cues perceptions of gender; Brown and Perrett 1993) come to be subjectively perceived by the brain as categorical. Are there separable neural representations of both the sexually dimorphic signal objectively apparent in the face versus how that signal is subjectively perceived? Neural activity involved in processing a face's gender should be sensitive to the amount of sexually dimorphic content in the face. That is, brain regions that encode gender should show stronger responses as face stimuli become more sexually dimorphic (or, "gendered"), regardless of whether the stimuli are perceived to be male or female. We were thus interested in identifying brain regions whose responses to faces increase as a face becomes more gendered, both in terms of how gendered it *objectively* is and how gendered it *subjectively* appears to be.

We generated highly realistic faces that varied by apparent gender (–6 = extremely masculine, 0 = androgynous, 6 = extremely feminine, converted into absolute values for analysis: 0 = androgynous and 6 = extremely gendered; Fig. 1). Morphing allowed us to manipulate sexually dimorphic content and unconfound all other perceptual information. Measuring blood oxygenation level–dependent (BOLD) signals using functional magnetic resonance imaging (fMRI), 32 volunteers (16 males), passively viewed these faces each for 2 s in randomized order (see Experimental Procedure). Participants also completed an FFA localizer task using separate face stimuli. After scanning, participants made subjective dichotomous gender judgments (male or female) and gave subjective ratings of gender using a 9-point continuous scale (1 = extremely masculine, 5 = androgynous, 9 = extremely feminine).

Materials and Methods

Participants

Thirty-two right-handed healthy volunteers (16 males) participated in the study. All participants gave informed consent in a manner approved the Massachusetts Institute of Technology's Committee On the Use of Humans as Experimental Subjects and were paid for their participation. Participants were scanned at the Athinoula A. Martinos Center for Biomedical Imaging in Cambridge, MA.

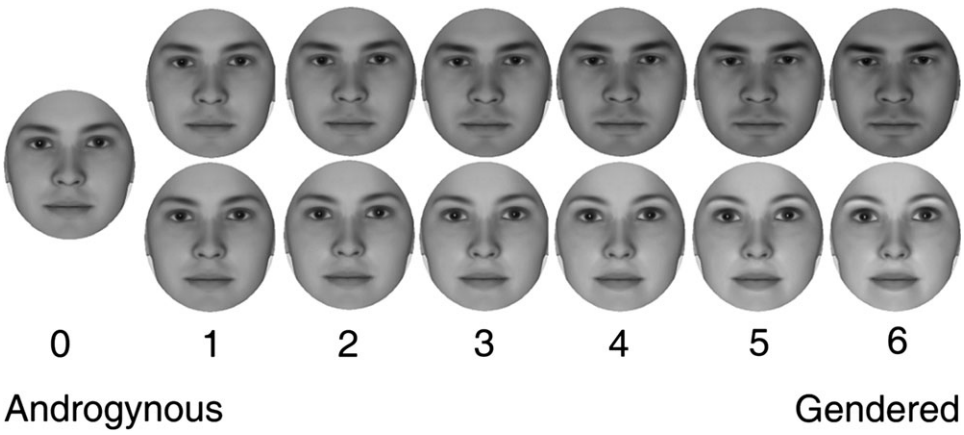

Figure 1. Sample set of face gender stimuli. Example of 1 set of morph stimuli (out of 8). Each set comprised of 13 morphs that varied by sexually dimorphic content (ranging from extremely male to extremely female), originating from one unique face identity.

Materials

We used FaceGen Modeler 3.1 (Singular Inversions) to generate highly realistic 3D faces that were seamlessly morphed by gender (from extremely male to extremely female), as based on anthropometric parameters of human population. Computational details on the morphing algorithm are provided by Blanz and Vetter (1999). We semirandomly generated 8 unique face identities. For each unique identity, we generated 13 face morphs that varied monotonically by gender (−6 = extremely male, 0 = androgynous, 6 = extremely female). Because all face stimuli were synthetically generated and created at some point along a continuum of sexually dimorphic content (rather than based on a real face which was subsequently morphed), no face stimuli appeared more morphed or synthetic than others. Face stimuli were presented frontally and cropped (see Fig. 1). For fMRI analyses, original morph values were converted into absolute values: 0 = androgynous and 6 = extremely gendered.

Experimental Procedure

The 104 face stimuli and an additional 52 baseline trials (fixation cross) were passively viewed each for 2 s in 1 of 3 pseudorandomized orders, each sequenced in a manner as to maximize the efficiency of event-related BOLD signal estimation (Dale 1999). Either prior to or following the primary scan (counterbalanced across participants), participants engaged in an FFA localizer task. For this task, participants viewed 116 faces (different than the ones used in the primary scan), 58 nonface objects, and 58 baseline trials (fixation cross) each for 2 s in 1 of 4 pseudorandomized orders, each sequenced in a manner as to maximize the efficiency of event-related BOLD signal estimation (Dale 1999). Following the scan, participants made dichotomous gender judgments (male or female) of face stimuli and gave ratings along a 9-point continuous scale (extremely masculine to extremely feminine).

fMRI Acquisition

Participants were scanned using a Siemens 3T Tim Trio scanner at the Massachusetts Institute of Technology, Martinos Imaging Center. Anatomical images were acquired using a T_1-weighted protocol (256 × 256 matrix, 128 × 1.33 mm sagittal slices). Functional images were acquired using a single-shot gradient echo-planar imaging sequence (time repetition [TR] = 2000 ms, time echo = 30 ms). Thirty-two interleaved oblique axial slices (3.125 × 3.125 × 5 mm voxels; slice gap = 1 mm) parallel to the anterior commissure-posterior commissure line were obtained. Analysis of the imaging data was conducted using BrainVoyagerQX (Brain Innovation, Maastricht, The Netherlands). Functional imaging data preprocessing included 3D motion correction, slice scan time correction (sinc interpolation), spatial smoothing using a 3D Gaussian filter (7-mm full width at half-maximum), and voxelwise linear detrending and high-pass filtering of frequencies (above 3 cycles per time course). Structural and functional data of each participant were transformed to standard Talairach stereotaxic space.

FFA Localization Analysis

We modeled BOLD responses during the FFA localizer task in an event-related design using a general linear model (GLM), with the face and object conditions modeled as boxcar functions convolved with a 2-gamma hemodynamic response function (HRF). To localize individual participants' FFA, we conducted a faces > objects contrast using a voxelwise threshold of $P < 0.0001$ in each participant. Functional regions of interest (ROIs) of participants' FFA were defined as voxels in the right fusiform gyrus (FG) that were elicited by this contrast. We ensured that voxels in the bilateral occipital cortex elicited by this analysis were not included in FFA localization as these were likely to denote the occipital face area, not the FFA. If no cluster was found in the right FG, we gradually relaxed threshold until a cluster emerged (most liberal: $P < 0.01$), consistent with previous work. The FFA could not be localized in 3 participants, leaving 29 participants for the ROI analysis.

Face Gender Analysis

To identify neural representations of objective face gender, original morph values of face stimuli (−6 = extremely male, 0 = androgynous, 6 = extremely female) were converted into absolute values (0 = androgynous, 6 = extremely gendered). A parametric predictor of objective gender was created, which modeled the corresponding objective gender value at the onset of each face stimulus. Individual participants' BOLD signals were modeled in an event-related design using a GLM with 2 orthogonal predictors: one coding all face stimuli and the separate parametric predictor of objective gender, described above, which was z-normalized within participants. This predictor coded the amount of objective gendered content viewed by the participant for each face stimulus. To identify neural representations of subjective face gender, participants' postscan continuous gender ratings (1 = extremely male, 5 = androgynous, 9 = extremely female) were subtracted by a constant of 5 and converted into absolute values. A parametric predictor of subjective gender was created, which modeled the corresponding subjective gender value at the onset of each face stimulus. For each participant, this corresponding subjective gender value was determined by that same participant's idiosyncratic judgment of that same face stimulus (as indicated postscan). In another event-related GLM design matrix, BOLD signals were modeled with 2 orthogonal predictors: one coding all face stimuli and one separate parametric predictor of subjective gender, described above, which was z-normalized within participants. Thus, this predictor coded the amount of gendered content subjectively perceived by the participant for each face stimulus. In both GLM analyses, conditions were modeled as boxcar functions (for parametric predictors, the amplitude of which was parametrically varied) and convolved with a 2-gamma HRF. All first-level GLM analyses conducted on individual participants' fMRI signal were submitted to a second-level random effects analysis, treating participants as a random factor. Statistical maps depicting BOLD responses were overlaid onto an average image of all participants' corresponding structural data.

We conducted an ROI analysis of the FFA. To determine whether the FFA was reliably modulated by objective gender or subjective gender, we contrasted these parametric predictors (described above) against baseline (in separate analyses). Parameter estimates of the magnitude of BOLD response (beta values) associated with the objective and subjective gender predictors were extracted from all voxels within participant-specific ROIs and compared against baseline (0) using 1-sample t-tests. We also conducted group-level whole-brain analyses testing the parametric effects of objective and subjective gender. To test these, separate contrasts were conducted to compare each parametric predictor against baseline (objective gender > baseline; subjective gender > baseline) using a voxelwise threshold of $P < 0.001$. Additionally, we determined whether regions elicited by these analyses survived under a more stringent false discovery rate (FDR)–corrected threshold (Genovese et al. 2002).

To further explore the parametric effect of objective face gender in regions elicited by the whole-brain analysis, we constructed another GLM design matrix which included a total of 7 predictors used to separate out BOLD signals associated with each level of objective face gender: face (at morph level 0), face (at morph level 1), face (at morph level 2), face (at morph level 3), face (at morph level 4), face (at morph level 5), and face (at morph level 6). Thus, each predictor was coded as 1 at the onset of face stimuli corresponding with that objective morph level and as 0 at the onset of face stimuli not corresponding with that morph level. We extracted beta values, averaged across all voxels within regions elicited by the whole-brain analysis (testing the parametric effect of objective gender), that corresponded with each of the 7 predictors.

To investigate the possibility that an fMRI adaptation effect spuriously produced the results in fusiform cortex, we constructed another GLM design matrix that included a total of 14 predictors used to separate out BOLD signals associated with each level of objective gender during the first half of the experiment (face onsets during the first 79 TRs) versus the second half (face onsets during the last 78 TRs).

All reported t-tests are 2-tailed.

Results

Behavioral Results

Participants' postscan dichotomous gender judgments were coded as male = 0 and female = 1. Ratings of continuous gender

were rescaled to vary between 0 (extremely masculine) and 1 (extremely feminine). To determine the relationship between objective morph values and subjective perceptions, we used generalized linear model polynomial regression analyses using a generalized estimating equation (GEE) approach (Zeger and Liang 1986). We adopted this approach because our design involved repeated measurements (whose intracorrelations needed to be appropriately accounted for), and one response variable was binary, whereas the other was continuous. This approach allows the linear model to be related to the response variable via a link function, which can be in the form of a logit function (e.g., for binary data) or identity function (e.g., for continuous data). Thus, GEE can handle both normal and logistic regression and account for the intracorrelations in a repeated-measures design, making it uniquely appropriate for the present analyses. Moreover, it is able to simultaneously incorporate trial-by-trial data from every participant. We report unstandardized regression coefficients.

We regressed subjective dichotomous gender judgments onto the linear component of objective morph values, in addition to quadratic and cubic components (using logistic regression). As a face's objective morph value rose from extremely male (–6) to extremely female (6), it was more likely to be perceived as female, as indicated by a significant linear component, $B = 1.641$, standard error (SE) = 0.102, $P < 0.0001$. This analysis also elicited a significant cubic component ($B = -0.012$, SE = 0.006, $P < 0.05$), which captured a step-like sigmoidal function, consistent with a categorical perception effect (Fig. 2*A*).

It is possible that the observed nonlinear relationship between objective and subjective face gender may have been an artifact of participants being constrained to making dichotomous gender judgments (see Campanella et al. 2001). Thus, stronger evidence of a categorical perception effect could be obtained by finding a nonlinear relationship between objective and subjective face gender when participants used the more sensitive 9-point continuous scale. We regressed subjective continuous gender ratings onto linear, quadratic, and cubic components of objective morph values (using normal regression). As a face became more objectively female, it was perceived to be more female, as indicated by a significant linear component, $B = 0.182$, SE = 0.006, $P < 0.0001$. This analysis also indicated the presence of highly significant quadratic ($B = -0.002$, SE < 0.001, $P < 0.01$) and cubic ($B = -0.002$, SE < 0.001, $P < 0.0001$) components. As shown in Figure 2*B*, morph values had a strong influence on continuous gender ratings near the category boundary (shown in Fig. 2*A*; ca. ranging between morph values –1 and 2), indicating by the sharp slope, but this influence was attenuated on either side of the boundary. This is consistent with the phenomenon of categorical perception (Harnad 1987).

Thus, assessed both categorically and continuously, subjective perceptions of face gender exhibited strong nonlinear effects despite linear monotonic changes in the objective perceptual signal, confirming that the gender of these faces was perceived in categorical fashion. As we were interested in brain responses to the level of sexually dimorphic content, regardless of whether faces were perceived to be male or female, we plotted absolute values of subjective ratings of gender (androgynous to extremely gendered) as a function of the objective amount of sexually dimorphic content (Fig. 2*C*). Shown in the figure, the nonlinear rise of this plotted curve (describing subjective perceptions of sexually dimorphic content) is distinctly warped from the linear function that maps onto the objective amount apparent in the face. This suggests that subjective perceptions were transformed in a manner consistent with the categorical perception phenomenon, where differences in faces closer to the gender category boundary (morph values 0–3 in Fig. 2*C*) are amplified (indicated by the steeper slope), whereas differences in faces further from the boundary (morph values 4–6 in Fig. 2*C*) are attenuated (indicated by the shallower slope). Our neuroimaging analyses examined how the brain differentially tracks these 2 encoding functions: objective parameters of sexually dimorphic content versus subjective perceptions of it.

Neuroimaging Results

The right FFA was of a priori interest given its role in face perception (Kanwisher and Yovel 2006). ROIs of the FFA were defined on an individual basis using voxels in the right FG that were elicited by a face > objects contrast. The FFA could not be localized in 3 participants, leaving a total of 29 participants for

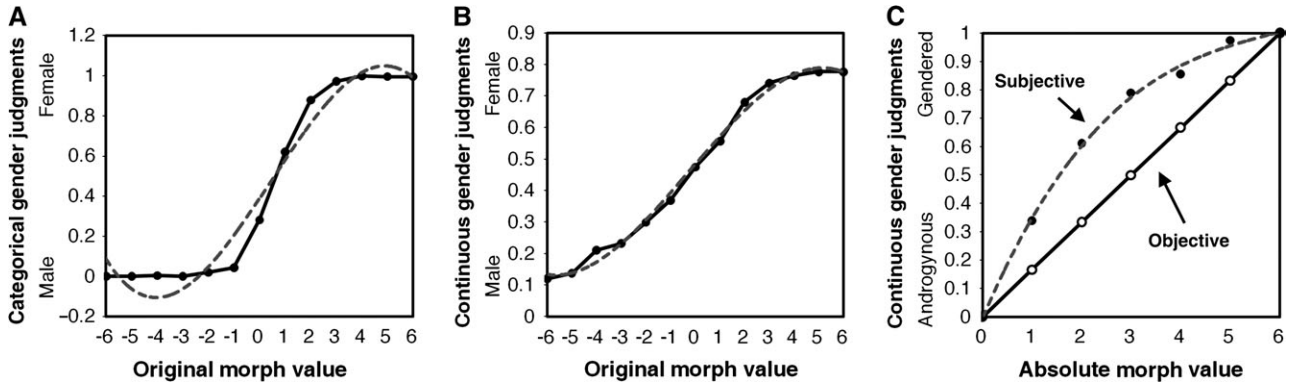

Figure 2. Behavioral categorical perception effects. (*A*) Subjective categorical gender judgments are plotted as a function of morph values, showing nonlinear effects in the form of a step-like sigmoidal function. (*B*) Subjective gender judgments made on a continuous scale (standardized to vary between 0 and 1) are plotted as a function of morph values, showing nonlinear effects where changes along the continuum are amplified near the category boundary. (*C*) Absolute values of standardized continuous gender judgments are plotted as a function of the absolute value of morph values, depicting a nonlinear curve associated with subjective gender encoding. Also depicted is a linear function associated with objective gender parameters. For comparison, subjective judgments and objective parameters were both fit to range between 0 and 1. This reflects the warping of subjective gender perceptions away from the objective signal due to categorical perception effects.

ROI analysis. We examined whether FFA responses were reliably modulated by objective gender, subjective gender, or both. Beta values associated with the objective and subjective gender predictors were extracted from all voxels within participant-specific ROIs and compared against baseline (0) using 1-sample *t*-tests. These revealed that, as the amount of objective sexually dimorphic content monotonically increased, BOLD responses in the FFA linearly increased, 1-sample $t_{28} = 2.40$, $P < 0.05$. Dissimilarly, the FFA was not modulated by subjective perceptions of sexually dimorphic content, 1-sample $t_{28} = 0.81$, $P = 0.42$. For comparison, these t values, indexing the strength of the correlations between FFA BOLD response and the objective and subjective predictors, appear in Figure 3.

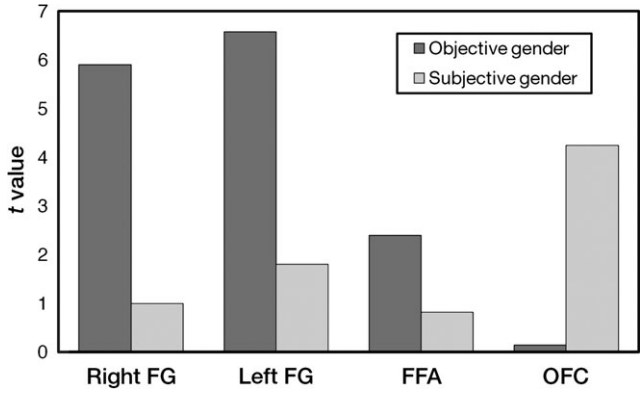

Figure 3. Encoding of objective and subjective gender. Strength of the correlations (*t* values) with the parametric models of objective and subjective gender. The right FG, left FG, and FFA exhibited responses positively correlating with objective parameters, but these were substantially less correlated with subjective perceptions. Conversely, the OFC exhibited responses positively correlating with subjective perceptions, but these were substantially less correlated with objective parameters. There are no error bars to plot because *t* values are depicted.

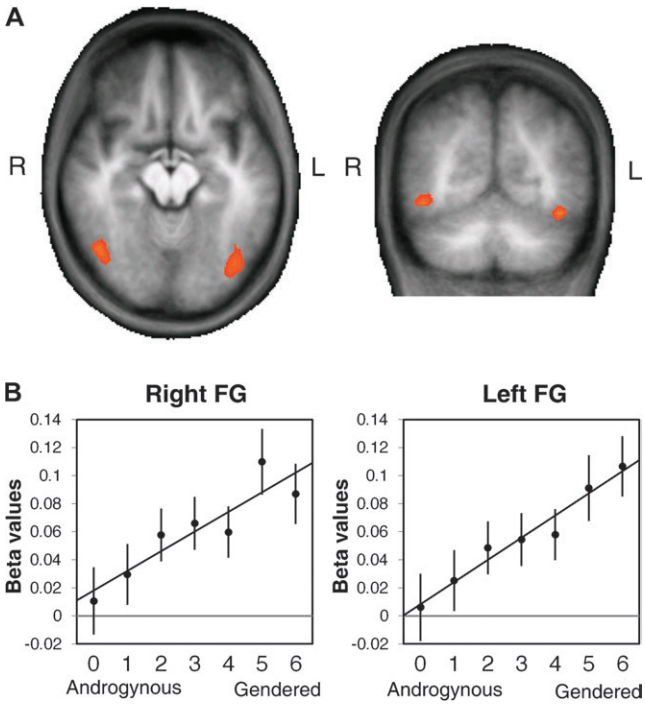

Figure 4. Parametric effect of objective face gender. (*A*) Whole-brain analysis testing a parametric modulation by objective gender, revealing responses in a bilateral region in lateral FG (Fig. 3 and Table 1). (*B*) The magnitude of response (beta values) in the lateral FG is plotted as a function of each level of objective gender. As sexually dimorphic face content monotonically increased, this region showed linearly increasing responses. Error bars denote standard error of the mean.

To identify neural representations of objective face gender beyond the FFA, we conducted a whole-brain analysis contrasting the parametric effect of objective gender against baseline ($P < 0.001$, uncorrected). This analysis revealed that activity in a lateral region of the FG (BA 19/37), bilaterally, was associated with objective linear changes in face gender (Fig. 4*A* and Table 1). This parametric modulation was robust, surviving at a more stringent FDR-corrected threshold ($P < 0.05$, corrected). To directly compare the strength of correlation between BOLD responses in this region and the objective versus subjective gender predictors, beta values associated with each predictor were extracted from all voxels within this region (separately for the left and right hemispheres). One-sample *t*-tests comparing these beta values to baseline (0) revealed that lateral FG responses correlated better with objective gender (right: $t = 5.90$; left: $t = 6.57$) than subjective gender (right: $t = 0.10$; left: $t = 1.80$), depicted in Figure 3. To better specify the nature of this region's modulation by objective gender, we extracted beta values to face stimuli at each separate level of objective gender (see Materials and Methods). As the objective

Table 1
Regions of activation elicited by the whole-brain analysis testing the parametric effect of objective face gender (FDR, $P < 0.05$, corrected) and subjective face gender ($P < 0.001$, uncorrected)

Region	Side	*x*	*y*	*z*	mm^3
Objective face gender					
Fusiform gyrus	L	−37	−67	−10	2493
Fusiform gyrus/middle temporal gyrus	R	42	−61	−4	1361
Middle occipital gyrus	R	30	−78	18	243
Middle occipital gyrus	L	−30	−86	−6	51
Subjective face gender					
OFC	M	−1	53	−7	62

Note: L, left; R, right; and M, medial.

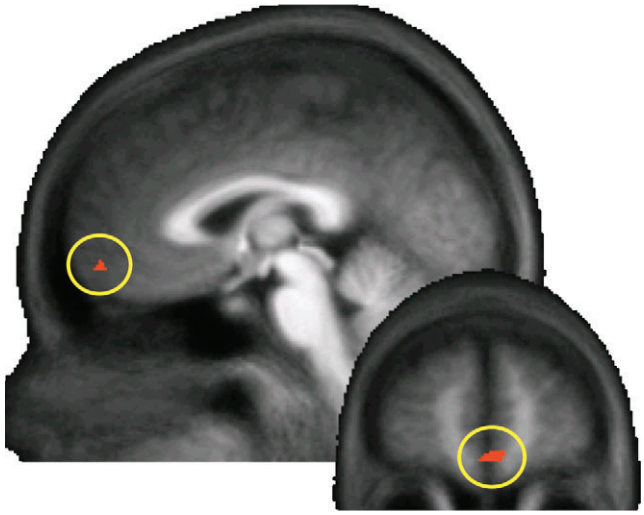

Figure 5. Parametric effect of subjective face gender. (*A*) Whole-brain analysis testing a parametric modulation by subjective gender, revealing responses in the OFC, which showed responses that positively correlated with subjective gender (Fig. 3).

amount of sexually dimorphic content monotonically increased, this lateral FG region showed linearly increasing responses (Fig. 4*B*). Interestingly, 1-sample *t*-tests comparing the beta values of each level of objective gender with 0 (baseline) revealed that the most androgynous faces (morph values 0 and 1) did not significantly engage this gender-sensitive FG region relative to baseline: right (1-sample t_{31} values = 0.31 and 1.03, *P* values = 0.76 and 0.31, respectively) and left (1-sample t_{31} values = 0.44 and 1.36, *P* values = 0.66 and 0.18, respectively). Thus, this region's sensitivity to objective face gender does not reflect some overall selectivity to faces. Only when a face began to depict a gender did this region activate to encode it (starting with morph value 2; left: 1-sample t_{31} = 2.48, $P < 0.05$; right: 1-sample t_{31} = 3.04, $P < 0.01$) and did so in a graded fashion (see Fig. 4*B*). This underscores this region's particular sensitivity to objective face gender.

To identify neural representations of subjective perceptions of face gender, we conducted a whole-brain analysis contrasting the parametric effect of subjective gender against baseline ($P < 0.001$, uncorrected). This analysis revealed a localized region of the orbitofrontal cortex (OFC; Fig. 5 and Table 1). Beta values associated with the subjective and objective gender predictors were extracted from all voxels within this region. One-sample *t*-tests comparing these beta values to baseline (0) indicated that OFC correlated better with subjective gender ($t = 4.24$) than objective gender ($t = 0.15$), depicted in Figure 3. The positive mean *t* value associated with this region's modulation by subjective gender indicates that, as a participant's subjective perception of gendered content increased, this OFC region sensitively encoded it by exhibiting stronger responses.

An alternative interpretation of the parametric modulation of objective face gender in the lateral FG and FFA is that, rather than responding to objective gender parameters, the modulation simply reflects an fMRI adaptation effect. fMRI adaptation refers to the phenomenon whereby repeated presentation of visual stimuli, such as a face, results in the reduction of BOLD responses in regions involved in visuoperceptual processing, such as face-responsive areas in fusiform cortex (Grill-Spector and Sayres 2008, for review). This reduction is most pronounced for repetitions of the exact same stimulus, but modified versions of a stimulus can also result in BOLD reduction. Thus, it could be that the lateral FG and FFA show stronger responses to more gendered faces (and weaker responses to more androgynous faces) because, as a face becomes more androgynous, it approximates the composite of all morph variants within that facial identity. If the case, more androgynous faces might have elicited weaker mean BOLD responses because the lateral FG and FFA experienced more adaptation to them (as they reflect the greatest overlap among morph variants and this overlap would have been repeated more).

To ensure that such fMRI adaptation effects were not spuriously driving the effects here, we compared BOLD responses associated with each level of objective face gender during the first half of the experiment to the second half (see Materials and Methods). If the lateral FG's and FFA's modulation by objective gender was due to fMRI adaptation, BOLD responses in these regions would be reduced in the second half of the experiment when compared with the first half (because more face repetitions would have occurred), and this would be especially true for the more androgynous faces. For each level of objective gender, left and right lateral FG and FFA BOLD responses during the first and second halves of the experiment did not significantly differ: morph value 0 (*P* values = 0.64, 0.58, and 0.60, respectively), morph value 1 (*P* values = 0.12, 0.12, and 0.87, respectively), morph value 2 (*P* values = 0.56, 0.22, and 0.24, respectively), morph value 3 (*P* values = 0.88, 0.84, and 0.92, respectively), morph value 4 (*P* values = 0.06, 0.58, and 0.49, respectively), morph value 5 (*P* values = 0.25, 0.17, and 0.15, respectively), and morph value 6 (*P* values = 0.59, 0.78, and 0.64, respectively). Also note that the *P* values exhibit no clear trend of lower morph values (more androgynous faces) being closer to reaching significance than higher morph values (more gendered faces), which would be expected if adaptation were driving the effects here. This analysis alleviates the concern that the lateral FG's and FFA's modulation by objective gender was produced by fMRI adaptation.

Linear Responses to Objective Gender in Fusiform Cortex

It is possible that the linear and graded pattern of responding in the FFA and lateral FG to objective sexually dimorphic content, reported above, could have been spuriously produced by averaging together different forms of nonlinear (categorical) step functions in each individual participant. If fusiform responses' "step" occurred at different levels of objective gender across participants, when averaged together, mean beta values could feign gradiency and linearity (as shown in Fig. 4*B*), when in reality, strong nonlinear step functions that varied across participants would underlie the effect. To eliminate this possibility, 2 additional analyses were conducted.

First, we inspected the distributions of beta values in the FFA and lateral FG. If fusiform BOLD responses in each individual participant were distributed as only primarily 2 values belonging to a categorical step function (i.e., nongendered vs. gendered), when beta values are pooled across participants, their distribution would exhibit bimodality due to values gravitating toward 2 separate means. Bimodality may be tested using the bimodality coefficient (*b*), which has a standard cutoff of $b > 0.555$ (SAS Institute 1989). If *b* exceeds this cutoff, unimodality is rejected in favor of bimodality. Histograms of the beta value distributions from the FFA, right FG, and left FG (pooling across all levels of objective gender and all participants) appear in Figure 6. The FFA distribution (skewness = –0.073, kurtosis = –0.084), right FG distribution (skewness = 0.050, kurtosis = –0.548), and left FG distribution (skewness = –0.048, kurtosis = –0.189) showed no evidence of bimodality: *b* values = 0.340, 0.402, 0.351, respectively. Moreover, Shapiro–Wilk tests confirmed that these distributions did not significantly depart from a normal distribution (*P* values = 0.67, 0.32, and 0.80, respectively), eliminating the possibility that they hosted latent bimodal features.

Second, we compared the strength of fusiform regions' correlation with the linear function of objective gender versus their correlation with a simulated nonlinear step function. Our aim was to demonstrate that fusiform responses were predicted better by a linear model of objective gender (representing gender's inherent gradiency) than by a nonlinear model involving a step function (lacking any graded representation). We constructed 2 additional GLM design matrices to model 2 variants of a nonlinear step function potentially underlying fusiform responses. Each design matrix included 2 orthogonal predictors: one for all face stimuli and

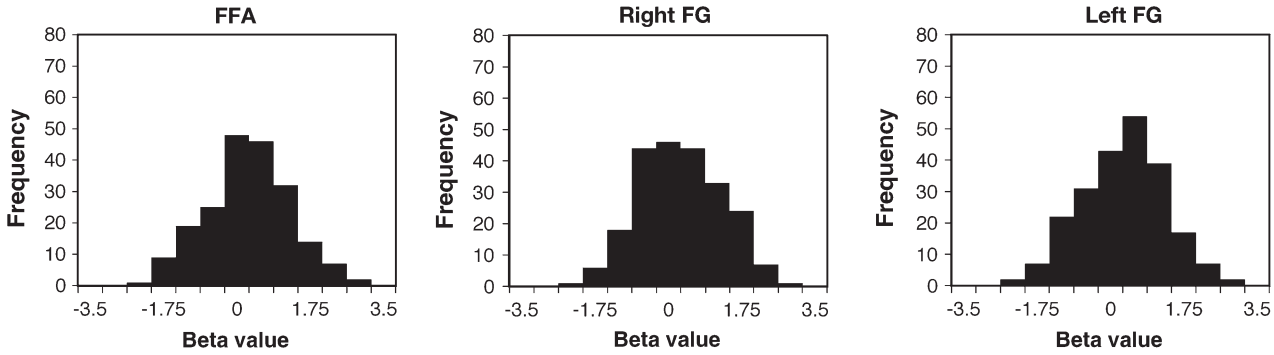

Figure 6. Distribution of fusiform responses to objective face gender. Histograms depicting *z*-normalized distributions of beta values (pooled across all 7 levels of objective face gender and all participants) in the FFA, right FG, and left FG. The plots illustrate normality and a lack of bimodality.

a separate parametric predictor that modeled the step function. For the first model, the parametric predictor was specified by recoding objective gender values [0, 1, 2, 3, 4, 5, 6] into [0, 0, 0, 0, 1, 1, 1]; for the second model, they were recoded into [0, 0, 0, 1, 1, 1, 1]. That is, the parametric predictor of the first model coded faces with objective morph values 0–3 as 0 (nongendered) and faces with morph values 4–6 as 1 (gendered); that of the second model coded faces with morph values 0–2 as 0 (nongendered) and faces with morph values 3–6 as 1 (gendered). Thus, these predictors model 2 possible variants of a categorical step function that may spuriously underlie fusiform responses. Beta values associated with these 2 nonlinear step predictors were extracted from the FFA, left FG, and right FG. Separate 1-sample *t*-tests comparing these to 0 (baseline) revealed that the FFA correlated better with the linear and graded model of objective gender ($t = 2.40$; reported earlier) than either nonlinear step model (t values = 1.28 and 1.16, respectively). The right and left FG regions also correlated better with the linear model (right: $t = 5.90$; left: $t = 6.57$; reported earlier) than the nonlinear models (right: t values = 2.85 and 2.90, respectively; left: t values = 4.24 and 4.13, respectively).

Taken together, these analyses cast doubt on the possibility that nonlinear categorical responses varying across participants spuriously produced an average effect of linearity in fusiform BOLD signals, cementing our evidence for linear graded responses in fusiform cortex to gendered face content.

Discussion

These results provide evidence for graded neural representations of face gender in the FFA, lateral FG, and OFC. We show that these regions, independent of attention, encode the gender of another's face. The FFA and lateral FG were involved in representing information about objective monotonic linear changes in apparent sexually dimorphic face content, whereas the OFC was involved in representing the amount of sexually dimorphic content that was subjectively perceived. Behaviorally, we found that subjective perceptions of face gender, represented in the OFC, had a strong nonlinear relationship with objective face gender in a manner suggesting that they had been transformed by categorical perception effects (Harnad 1987; Campanella et al. 2001), as shown in Figure 2*C*.

Although face gender was perceived categorically, consistent with previous work (Campanella et al. 2001), we found that the FFA and lateral FG nevertheless showed patterns of responding that were linearly related to graded and objective perceptual information and not subjective perceptions (which were warped and nonlinearized and represented in the OFC). Thus, despite the categorical warping of subjective perceptions, we find that the objective gradiency of gender inherent in the face is linearly tracked in relatively early regions of face-related visuoperceptual processing: the FFA and lateral FG.

This evidence is consistent with the point of view that categorical perception effects result from a gradual competition process (e.g., for face gender, between "Male" and "Female"), where objective gradiency is initially represented in neuronal populations, which then warps over time into subjective nonlinearized perceptions (e.g., McMurray and Spivey 2000; McMurray et al. 2003; Spivey 2007). Our findings suggest that the brain does indeed represent the objective gradiency of face gender in such initial perceptual representations and does so in the FFA and lateral FG. These fusiform representations may then be subsequently transformed during this competition process, the results of which are transmitted to the OFC (which we found to correlate with such transformed subjective perceptions). This is consistent with our finding that the OFC correlated better with subjective perceptions than objective parameters, but fusiform regions correlated better with objective parameters than subjective perceptions (Fig. 3).

The transmission of FG representations of objective face gender parameters to the OFC, which represents their subjective warping, is likely considering that fusiform cortex sends major visuoperceptual output directly to (and also receives information back from) the OFC (Carmichael and Price 1995). From neurophysiological work, it has long been known that there is a population of face-responsive neurons in OFC (Thorpe et al. 1983). These neurons have firing rates that can be remarkably similar to those in visuoperceptual temporal cortex (such as lateral FG), tending to respond with later latencies (ca. 130–220 ms) than those in temporal regions (ca. 80–100 ms; Rolls 2000). This suggests that results from temporal cortical face processing, such as FG representations of objective gender, are projected up to this face-responsive neuronal population in OFC. The OFC been implicated in processing facial emotional cues, with emotion recognition impaired by OFC lesions (Hornak et al. 1996) and by transient OFC disruption in healthy participants using transcranial magnetic stimulation (Harmer et al. 2001). Here we extend

these findings by showing that OFC responses represent another face cue: subjectively perceived sexually dimorphic content. Importantly, visual neurons in OFC have been shown to reflect learned properties of visual stimuli rather than actual physical properties, which distinguishes them from those in temporal cortex (Rolls 2000). It is thus fitting that we find the OFC to encode subjective parameters of gender, whereas it did not encode objective physical parameters of gender (Fig. 3). Moreover, recent evidence suggests that fast magnocellular projections from visual cortex to the OFC, in combination with a top-down projection from the OFC to fusiform cortex, play a critical role in top-down facilitation of visual object recognition (Kveraga et al. 2007). Thus, such top-down facilitation mediated by the OFC may be important for biasing objective gender representations available in fusiform cortex toward subjective categorically warped perceptions.

OFC damage also leads to impairments in recognizing emotional expressions in the voice channel (Hornak et al. 1996), and given the high level of multimodal sensory integration in OFC (Rolls 2000), perhaps face-triggered representations in this region code higher-order social information linked to faces, such as how masculine or feminine a person is, which is based on subjectively perceived gendered face content. These OFC representations could also track other consequences of face categorization. For instance, the OFC plays a role in reward association learning (e.g., Gottfried et al. 2003), and thus, representations of subjective gender in the OFC could potentially be a product of the perceptual fluency, pleasantness, or ease with which face gender is implicitly perceived. These representations might also track other characteristics about the face stimuli that were rewarding. Future work could further characterize the particular role of the OFC in categorical face perception.

We should note that our finding of the FFA's and lateral FG's encoding of graded objective parameters of the face, such as sexually dimorphic content, is consistent with evidence showing that representations in the face perceptual system are structured in a multidimensional face-space, where the initial encoding of physical face parameters would be highly linear, as found here (Tanaka et al. 1998). In this face-space, "Male" and "Female" would function as attractors that compete to pull neural face responses into settling onto a stable representation of gender (Tanaka et al. 1998; Campanella et al. 2001). Recent work shows that such attractor dynamics of a multidimensional face-space are evident in real-time behavioral markers that attempt to capture such dynamics online (Spivey 2007) and specifically with face gender (Freeman et al. 2008). Thus, fusiform regions may provide a version of this multidimensional face-space that linearly encodes graded parameters of the face, such as sexually dimorphic content (Freeman et al., forthcoming; also see Rotshtein et al. 2005).

These findings also extend evidence demonstrating that the lateral FG is involved in representations of static structural cues, such as identity, and is not involved (e.g., Haxby et al. 2000) or less involved (e.g., Calder and Young 2005) in dynamic expressions, such as emotional or speech cues. Although virtually all such work has examined the static cue of face identity, here we show that the lateral FG is also involved in representing another static cue: gender.

In sum, here we have identified attention-independent graded neural representations of face gender in the FFA and lateral FG, which track objective parameters, and OFC, which tracks subjective parameters. Face gender is perceived categorically, and we traced how this categorical perception is realized in the human brain. The results show that initial perceptual representations in FG and the FFA encode objective gender parameters, and these subsequently undergo transformations that become represented in the OFC, reflecting subjective perceptions.

Funding

National Science Foundation (# 0724416 to N.A.); a National Science Foundation graduate research fellowship to N.O.R.

Notes

We thank Jasmin Cloutier for helpful discussion. *Conflict of Interest*: None declared.

Address correspondence to Jonathan B. Freeman, Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155, USA. Email: jon.freeman@tufts.edu.

References

Beale JM, Keil CF. 1995. Categorical effects in the perception of faces. Cognition. 57:217–239.

Blanz V, Vetter T. 1999. A morphable model for the synthesis of 3D faces. In: SIGGRAPH'99; 1999 August. Los Angeles: ACM Press. p. 187–194.

Bornstein MH, Korda NO. 1984. Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. Psychol Res. 46:207–222.

Brown E, Perrett DI. 1993. What gives a face its gender? Perception. 22:829–840.

Calder AJ, Young AW. 2005. Understanding the recognition of facial identity and facial expression. Nat Rev Neurosci. 6:641–651.

Calder AJ, Young AW, Perrett DI, Etcoff NL, Rowland D. 1996. Categorical perception of morphed facial expressions. Vis Cogn. 3:81–117.

Campanella S, Chrysochoos A, Bruyer R. 2001. Categorical perception of facial gender information: behavioural evidence and the face-space metaphor. Vis Cogn. 8:237–262.

Carmichael ST, Price JL. 1995. Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. J Comp Neurol. 363:642–664.

Dale AM. 1999. Optimal experimental design for event-related fMRI. Hum Brain Mapp. 8:109–114.

Etcoff NL, Magee JJ. 1992. Categorical perception of facial expressions. Cognition. 44:227–240.

Fischer H, Sandblom J, Agneta H, Fransson P, Wright C, Backman L. 2004. Sex-differential brain activation during exposure to female and male faces. Neuroreport. 15:235–238.

Freeman JB, Ambady N, Holcomb PJ. 2009. The face-sensitive N170 encodes social category information. Neuroreport. doi: 10.1097/WNR.0b013e3283320d54.

Freeman JB, Ambady N, Rule NO, Johnson KL. 2008. Will a category cue attract you? Motor output reveals dynamic competition across person construal. J Exp Psychol Gen. 137:673–690.

Genovese CR, Lazar NA, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 15:870–878.

Gottfried JA, O'Doherty J, Dolan RJ. 2003. Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science. 301:1104–1107.

Grill-Spector K, Sayres R. 2008. Object recognition: insights in from advances in fMRI methods. Curr Dir Psychol Sci. 17:73–79.

Harmer CJ, Thilo KV, Rothwell JC, Goodwin GM. 2001. Transcranial magnetic simulation of medial-frontal cortex impairs the processing of angry facial expressions. Nat Neurosci. 4:17–18.

Harnad S. 1987. Categorical perception: the groundwork of cognition. Cambridge: Cambridge University Press.
Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. Trends Cogn Sci. 4:223–233.
Hornak J, Rolls ET, Wade D. 1996. Face and voice expressions identification in patients with emotional and behavioral changes followed ventral frontal lobe damage. Neuropsychologia. 34:247–261.
Kanwisher N, Yovel G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci. 361:2109–2128.
Kveraga K, Boshyan J, Bar M. 2007. Magnocellular projections as the trigger of top-down facilitation in recognition. J Neurosci. 27:13232–13240.
Liberman AM, Harris KS, Hoffman HS, Griffith BC. 1957. The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol. 53:368–385.
Macrae CN, Bodenhausen GV. 2000. Social cognition: thinking categorically about others. Annu Rev Psychol. 51:93–120.
McMurray B, Spivey M. 2000. The categorical perception of consonants: the interaction of learning and processing. Proc Chicago Linguist Soc. 34:205–220.
McMurray B, Tannenhaus MK, Aslin RN, Spivey MJ. 2003. Probabilistic constraint satisfaction at the lexical/phonetic interface: evidence for gradient effects of within-category VOT on lexical access. J Psycholinguist Res. 32:77–97.
Murphy GL. 2002. The big book of concepts. Cambridge (MA): MIT Press.
Paller KA, Ranganath C, Gonsalves B, LaBar KS, Parrish TB, Gitelman DR, Mesulam MM, Reber PJ. 2003. Neural correlates of person recognition. Learn Mem. 10:253–260.
Rolls ET. 2000. The orbitofrontal cortex and reward. Cerebral Cortex. 10:284–294.
Rosch EH. 1978. Principles of categorization. In: Rosch EH, Lloyd B, editors. Cognition and categorization. Hillsdale (NJ): Erlbaum. p. 27–48.
Rotshtein P, Henson RNA, Treves A, Driver J, Dolan RJ. 2005. Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat Neurosci. 8:107–113.
SAS Institute 1989. SAS/STAT user's guide. Cary (NC): SAS Institute.
Spivey MJ. 2007. The continuity of mind. New York: Oxford University Press.
Stevenage VS. 1998. Which twin are you? A demonstration of induced categorical perception of identical twin faces. Br J Psychol. 89:39–57.
Tanaka J, Giles M, Kremen S, Simon V. 1998. Mapping attractor fields in face space: the atypicality bias in face recognition. Cognition. 68:199–220.
Thorpe SJ, Rolls ET, Maddison S. 1983. Neuronal activity in the orbitofrontal cortex of the behaving monkey. Exp Brain Res. 49:93–115.
Valentine T. 1988. Upside-down faces: a review of the effect of inversion upon face recognition. Br J Psychol. 79:471–491.
Winston JS, Henson RNA, Fine-Goulden MR, Dolan RJ. 2004. fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. J Neurophysiol. 92:1830–1839.
Zeger SL, Liang KY. 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 42:121–130.