

ORIGINAL ARTICLE

From Parts to Identity: Invariance and Sensitivity of Face Representations to Different Face Halves

Stefano Anzellotti^{1,2} and Alfonso Caramazza^{1,2}

¹Cognitive Neuropsychology Laboratory, Harvard University, Cambridge, MA 02138, USA

and ²Center for Mind/Brain Sciences, University of Trento, Trento 38068, Italy

Address correspondence to Stefano Anzellotti, Cognitive Neuropsychology Laboratory, Harvard University, 33 Kirkland Street, Cambridge, 02138 MA, USA.
Email: anzellot@fas.harvard.edu

Abstract

Recognizing the identity of a face is computationally challenging, because it requires distinguishing between similar images depicting different people, while recognizing even very different images depicting a same person. Previous human fMRI studies investigated representations of face identity in the presence of changes in viewpoint and in expression. Despite the importance of holistic processing for face recognition, an investigation of representations of face identity across different face parts is missing. To fill this gap, we investigated representations of face identity and their invariance across different face halves. Information about face identity with invariance across changes in the face half was individuated in the right anterior temporal lobe, indicating this region as the most plausible candidate brain area for the representation of face identity. In a complementary analysis, information distinguishing between different face halves was found to decline along the posterior to anterior axis in the ventral stream.

Key words: anterior temporal lobe, compositionality, face identity, holistic processing, object recognition

Introduction

The ability to recognize a person's identity is crucial for our daily life and social interactions. The most commonly used source of information to recognize a person's identity is probably the visual appearance of the face. One way to study how the brain processes visual information to recognize the identity of a face is to begin by individuating brain regions where neural activity distinguishes between pairs of face tokens (specific images of faces) that depict different people. Regions that distinguish face tokens that depict different people can still lack information about whether 2 different face tokens (e.g., a front view and a profile) depict the same person. This information is crucial for the recognition of face identity. As a second step, then, it is possible to assess whether regions that distinguish between face tokens depicting different people also encode information about whether different face tokens, despite their differences, depict the same person (i.e., whether they encode invariant face representations). In the process of computing invariant face representations, some information that is irrelevant for the recognition of identity may be

discarded. Therefore, as a complementary analysis to the investigation of the invariance of face representations, it is important to determine whether regions that distinguish between face tokens depicting different people are also sensitive to identity-irrelevant properties (e.g., they distinguish a front view and a profile of the same face).

In addition to the crucial role of face recognition to guide our social interactions, the investigation of the neural mechanisms underlying the recognition of face identity is important in the broader context of visual object recognition. Many of the image transformations that we need to discount for recognizing the identity of a face (changes in viewpoint, occlusion, . . .) also occur for nonface objects, raising the possibility that similar computations could be used to achieve invariance across those transformations for both faces and other objects. Furthermore, some properties of face-responsive neurons, such as preference for mirror symmetrical views (Freiwald and Tsao 2010), are also observed in neurons responding to paper clips (Logothetis and Pauls 1995), suggesting that at least some of the computational

principles used in face recognition and object recognition are similar.

Recent studies investigating whether face representations in ventral stream regions are invariant and/or sensitive to identity-irrelevant properties considered changes in facial expression (Nestor et al. 2011) and viewpoint (Kietzmann et al. 2012; Anzellotti et al. 2013). Missing is an investigation of how identity is recognized starting from different partial views of a face. The ability to recognize faces from partial views is important when we need to recognize a face discounting occluders. Partial views of faces can lack information about the spatial relations between some face parts (e.g., the top half of a face lacks information about the distance between the eyes and the mouth). Therefore, they partly disrupt holistic processing, which is thought to be important in face processing (Young et al. 1987; Tanaka and Farah 1993). Despite this disruption, participants are able to recognize a person's identity better than chance from face parts (Sadr et al. 2003).

In the present study, we individuated regions that distinguish between face tokens depicting different people, and we investigated the invariance and sensitivity of face representations in these regions to different face halves (Fig. 1). Regions that distinguish between face tokens depicting different people were found both within the ventral stream and in parietal cortex. Within the ventral stream, the results showed a gradual decrease in sensitivity to specific view progressing from posterior to anterior regions. Functional connectivity analysis revealed 3 clusters of regions showing high functional connectivity within each cluster: occipitotemporal regions, parietal regions, and anterior temporal regions. Finally, the results showed that invariance across different partial views is captured in the right anterior temporal lobe.

Materials and Methods

Participants

Thirteen participants (age range 20–30 years, mean 24 years) were tested in this experiment. Their consent was obtained according to the Declaration of Helsinki (2007). The project was approved by the Human Subjects Committees at the University of Trento and Harvard University.

Stimuli

Face images of 3 individuals seen in front view were used as stimuli. Images were generated using 3D graphics software (DAZ-3D), in order to control the position of the faces within the image and to equate the color, texture, and illumination of the faces. This approach permits to control for differences in the low-level properties of the faces at a local level, which is especially important given the widespread effects of retinotopy in large portions of the ventral stream (Hemond et al. 2007). For each of the 3 individuals, 5 different face stimuli were used: an image of the whole face, 2 face halves cut along the vertical midline, and 2 face halves cut along the horizontal midline (Fig. 1). The face halves were positioned where they would be in the context of the whole face; therefore, for instance, the left half of the face was located in the right visual field, and the right half of the face was located in the left visual field. The use of half faces rather than occluded faces was motivated by a set of considerations. First, the presence of occluders could introduce information about the distance between parts of the occluders and face parts, potentially leading to inflated classification accuracy between the different identities. Second, the absence of occluders implies that a whole face is the sum of the face halves stimuli—this fact was exploited in the present

study to investigate the patterns of response to a whole face as a function of the responses to face parts (see MacEvoy and Epstein 2009 for a similar analysis). However, given the absence of occluders, caution should be exercised in generalizing the conclusions of this experiment to naturalistic viewing conditions.

To prevent classification of different face halves on the ground of difficulty effects, stimuli were selected such that the accuracy in recognizing identity from the 4 different face halves was matched. Recognition accuracy was tested in a pilot experiment on a separate set of participants. The identity recognition accuracies for the 4 different face halves were high, and within 2 standard deviations of each other (accuracies: 93.67%, 93.33%, 91.22%, 92.67%; standard deviations: 3.61%, 1.80%, 2.59%, 1.27% for the left, right, lower, and upper halves).

Experimental Design

Each participant was tested on 2 consecutive days. On the first day, participants completed a behavioral training session in which they learned to discriminate between the 3 individuals. During the training, participants were only exposed to whole faces. The training consisted of 2 parts: in the first part, lasting approximately 6 min, participants learned to associate a face with a number from 1 to 3, matching images of the faces with the number written under the image. In the second part, lasting approximately 40 min, participants saw the faces without the number, and were given up to 2 s to respond pressing the correct number from 1 to 3 on the keyboard of a computer. Immediately after the offset of the face image, participants were shown the correct number corresponding to the face. Responses produced after the presentation of the feedback image were not recorded. On the following day, participants were instructed to consider one of the 3 individuals as the “target” identity. Which individual was considered as the target was changed for each participant to rule out that any effects found occur only for a specific pair of faces. Inside the scanner, images of the 3 individuals were presented (including the whole face images and the 4 different halves). Each trial consisted of a face image (500 ms) followed by a fixation cross (1500 ms). Participants responded by pressing a button with the index finger of the right hand when the target identity was presented, and a different button with the middle finger of the right hand when any of the distractors was presented (irrespective of whether the image was a whole face or a half face). This task was chosen because it leads participants to focus their attention on the identity of the faces, and, at the same time, it requires the same behavioral response for 2 different identities—the 2 distractors. As a consequence, it is possible to investigate the differences between the responses to the distractors while avoiding the confound of identity by behavioral response. The experiment consisted of 4 runs. Each run contained 305 trials (275 face stimuli and 30 null events), for a duration of approximately 10 min. Of the 275 face stimuli, 55 were images of the target (11 repetitions for each of the 5 image types), and 220 were images of the distractors (22 repetitions for each of the 5 image types for each of the 2 distractor identities). The runs were preceded by a standard functional localizer task containing whole faces, houses, and scrambled images, which was not used in the present analyses. None of the face stimuli used in the localizer task was used in the subsequent runs.

Data Collection

The data were collected on a Bruker BioSpin MedSpec 4T at the Center for Mind/Brain Sciences (CIMEC) of the University of

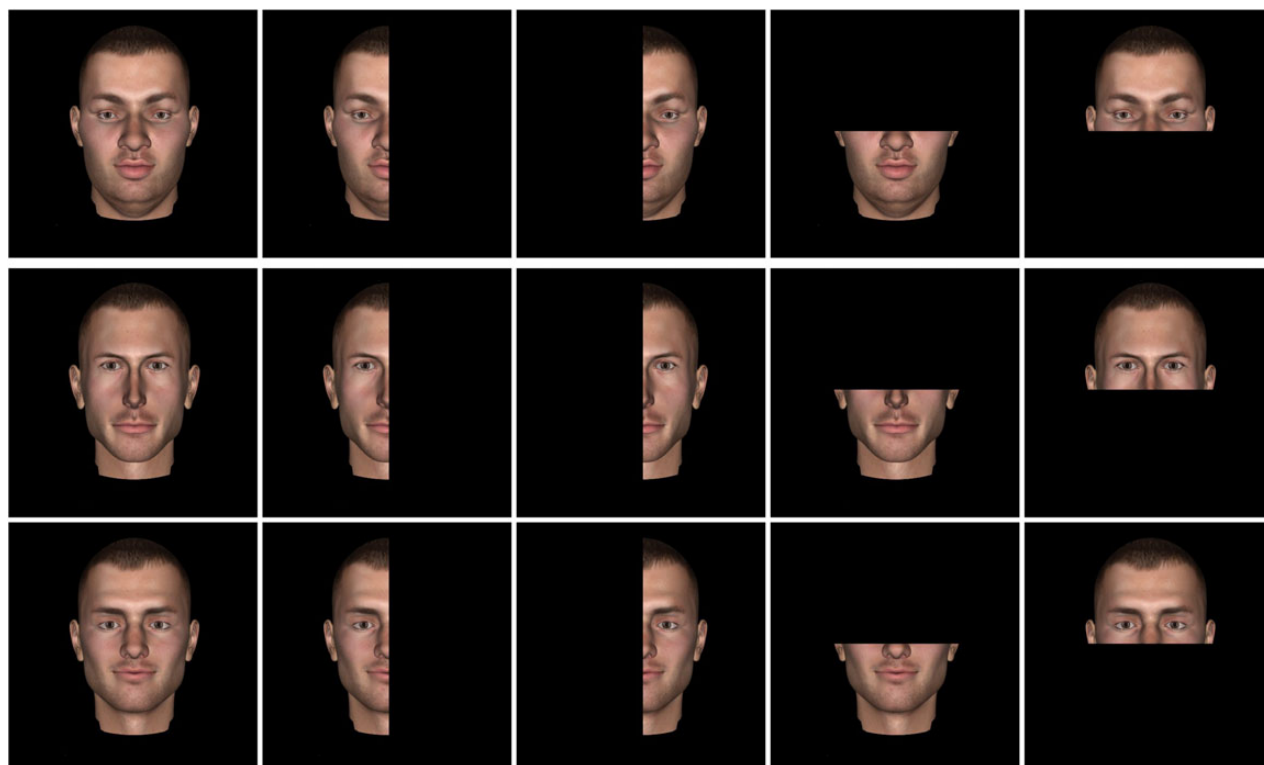


Figure 1. Images of the target faces (top row) and the distractors faces (bottom 2 rows). For each identity, images of the whole face and of 4 different face halves were used. Images were generated with DAZ-3D to control for low-level differences.

Trento using a USA Instruments eight-channel phased-array head coil. Before collecting functional data, a high-resolution ($1 \times 1 \times 1 \text{ mm}^3$) T_1 -weighted MPAGE sequence was performed (sagittal slice orientation, centric phase encoding, image matrix = 256×224 [Read \times Phase], field of view = $256 \times 224 \text{ mm}$ [Read \times Phase], 176 partitions with 1 mm thickness, GRAPPA acquisition with acceleration factor = 2, duration = 5.36 min, repetition time = 2700, echo time = 4.18, $TI = 1020 \text{ ms}$, 7° flip angle).

Functional data were collected using an echo-planar 2D imaging sequence with phase oversampling (image matrix = 70×64 , repetition time = 2000 ms, echo time = 21 ms, flip angle = 76° , slice thickness = 2 mm, gap = 0.30 mm, with $3 \times 3 \text{ mm}$ in plane resolution). Over 4 runs, 1260 volumes of 43 slices were acquired in the axial plane aligned along the long axis of the temporal lobe. Each run was preceded by a point-spread function sequence, and distortion correction was implemented with the method described by Zaitsev et al. (2004).

Data Analysis

Data were analyzed with SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>) and MARSBAR (Brett et al. 2002) running on MATLAB (2011a), and with custom MATLAB software using the MATLAB bioinformatics toolbox and LIBSVM (Chang and Lin 2011). Results were displayed with MRICron (Rorden and Brett 2000) and Caret (Van Essen et al. 2001).

Preprocessing

The first 4 volumes of each run were discarded and slice-acquisition delays were corrected using the middle slice as reference. All images were corrected for head movement and normalized to the standard SPM8 EPI template. The BOLD signal was high-pass

filtered at 128 s and prewhitened using an autoregressive model AR(1).

GLM

Data from the 4 runs were modeled with a standard general linear model (GLM) including motion regressors. One regressor was used for the target, and one for null events. In each run, all 10 distractor images were presented (2 identities by 5 image types), and every distractor image was repeated 22 times. The responses to each distractor image were modeled with 2 regressors: one for the first 11 presentations, and another for the remaining 11 presentations. This approach provides a reasonable balance between the accuracy of beta estimates and the number of data points available for training and testing in multivoxel pattern analyses (MVPAs).

Recursive Feature Elimination Mapping

Regions encoding information that distinguishes between face tokens depicting different people were identified with recursive feature elimination mapping (RFE mapping; De Martino et al. 2008) applied to the responses to whole faces. RFE mapping uses the weights of support vector machines (SVMs) to individuate the voxels that contribute most to a given classification. A SVM was trained to perform the classification between 2 faces using the data from all voxels in the gray matter in one experimental session. Each voxel was assigned a weight, which reflects the extent to which it contributes to the classification. The voxels with the smallest weights were removed, and the procedure was repeated in training a new SVM to perform the classification using only the remaining voxels. The fraction of voxels removed at each iteration decreased with exponential decay in subsequent iterations. The procedure completed in <10 steps.

This procedure was iterated until a fixed number of voxels was reached. The number of voxels was set to approximately half of the total number of voxels in the mask so as to maximize the number of possible combinations of voxel locations (given by the binomial coefficient). Selected voxels were assigned a value of 1, while nonselected voxels were assigned a value of 0. Minimal Gaussian smoothing was applied (4 mm full-width-at-half-maximum [FWHM]) in order to account for small differences across sessions and participants. The resulting maps for the different sessions and participants were averaged. This procedure yielded an average “frequency map” with values between 0 and 1 for each voxel. The value for a particular voxel indicates how frequently that voxel is individuated as being among the most informative for the particular classification tested, that is, how reliably that voxel contributes to the classification.

The significance of the reliability of the location of these informative voxels across sessions and participants was tested by generating statistical thresholds with Monte Carlo simulations. Simulated beta maps of the same dimension as the experimentally derived beta maps were generated for each session and participant. Simulating directly the BOLD time series would be complicated by their autocorrelated nature, which would call for the need to model the temporal autocorrelation structure of the real data and to recreate it in the simulated time series. The choice of simulating beta maps rather than BOLD time series removes (thanks to the randomized order of the experimental conditions) the need to explicitly model and recreate the temporal autocorrelation structure of the original BOLD time series. However, the issue of spatial correlation (smoothness) of the measure beta maps needs to be explicitly addressed. To this end, simulated beta maps were initially generated attributing to the different voxels instances of independent random variables with uniform distribution. Subsequently, the spatial smoothness of the real beta maps was estimated using SPM 8. This procedure yielded an estimated Gaussian kernel FWHM. The FWHM obtained from the real beta maps was then used to apply an identical amount of smoothing to the simulated time series, thus matching the spatial smoothness of the real and simulated beta maps. At this point, the simulated data were processed with the same RFE mapping algorithm applied to the real data. An SVM was trained to perform the same classification using simulated data extracted from all voxels in the same gray matter mask used for the real data. The same iterative procedure was used to select the most informative voxels. Importantly, the number of voxels selected in the simulated data for each session and participant was identical to the number of voxels selected in the real data. This procedure yielded maps of ones and zeros containing the same number of ones and zeros as the maps obtained for the real data. The same mild smoothing (4 mm FWHM) was applied to the maps obtained from the simulated data; therefore, the maps obtained from real and from simulated data were smoothed to the same degree. The maps for different sessions and participants were averaged, using an identical procedure as that used for the real data, yielding a “frequency map” as in the case of the real data. This procedure was iterated for 200 simulations of the data. For each of the 200 frequency maps obtained with this procedure, the maximum value in the map was stored in a vector, resulting in a 200-dimensional vector. Voxels in the frequency map obtained from the real data were considered to contribute significantly to classification if they had a value within the top 5% values in the 200-dimensional vector or higher. Note that this required “every” voxel in the frequency map obtained from the real data to have a frequency value equal or higher to the top 5% “whole-brain maxima” in the frequency maps

obtained from simulated data. In other words, using this threshold, in 95% of the frequency maps obtained from simulated data, zero voxels would be above threshold. This ensures that the threshold used is corrected for multiple comparisons.

Region-of-Interest Analysis: Support Vector Machines

The peaks obtained in the RFE mapping analysis were used as the centers of spherical regions of interest (ROIs) of radius 12 mm. Additional V1 ROIs matched for number of voxels were generated anatomically using the Wake Forest University PickAtlas toolbox for SPM (<http://fmri.wfubmc.edu/software/PickAtlas>). Within these ROIs, information about the face halves and about the identity of the faces was investigated. Since only the responses to the whole faces were used to individuate the regions, localization did not introduce any bias or circularity in the analysis. For each classification, a first-pass feature selection procedure was used to individuate the 300 voxels within the ROIs with the highest *F* values for the distinction between the conditions to be classified, using only the training data (to avoid circularity). SVMs with radial basis function kernel and flexible parameters *C* and *gamma* were used for classification. The parameters *C* and *gamma* were optimized using the training data only (again to avoid circularity), employing an exponential grid as recommended by the authors of the LIBSVM library (<http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>). To study information about specific face halves, a 4-way SVM classification was used. To study information about the identity of the face with tolerance across changes in half faces, a 2-way SVM classification was used on the 2 distractor faces, using 3 halves for training and the remaining fourth half for testing. Therefore, responses to different stimuli were used for the training and testing of the SVM. This procedure was iterated for all choices of the testing half, and the accuracy was averaged to obtain a value for each participant. The values obtained were then entered in a *t*-test to assess statistical significance.

Region-of-Interest Analysis: Pattern Reconstruction

To investigate the contributions of the response patterns to the left and right half faces to the response patterns to the whole faces in the ROIs, we modeled the pattern of response *W* to a whole face as a linear combination of the pattern of response *L* to the left half and the pattern of response *R* to the right half: $W = \alpha L + \beta R + E$, where *W*, *L*, *R* and *E* are vectors with a number of dimensions equal to the number of voxels in the ROI. The parameters α and β were obtained minimizing the cost function given by the sum of square errors across the different voxels (*E*²). The difference between the values of the parameters α and β was then tested with a *t*-test to investigate differences in the contribution of the left and right face halves to the response patterns to the whole faces. The same procedure was adopted to investigate the contributions of the response patterns to the top and bottom half faces to the response patterns to the whole faces.

Functional Connectivity Analysis

The relationship between the low-frequency fluctuations in the BOLD signal in different regions encoding information that distinguishes between face tokens depicting different people was investigated with functional connectivity. The preprocessed data were low-pass filtered (0.1-Hz cutoff frequency, see Van Dijk et al. 2010), and for each of 7 ROIs of 6 mm radius centered in the 7 peaks individuated with RFE mapping the time series in the different voxels within the ROI were averaged. The same procedure was applied to an additional 6-mm-radius control ROI in the right lateral ventricle (an average over a small sphere

was used to reduce potential effects of noise in the peak voxels). To remove fluctuations of no interest in the BOLD signal, the time series for the RFE ROIs were regressed on the time series for the control ROI and the residuals were considered for further analysis. Correlations between the residual time series were used to assess functional connectivity between the regions.

Results

Representations That Distinguish Between Face Tokens Depicting Different People

As a first step, brain regions that distinguish between face tokens depicting different people were localized with RFE mapping based on the responses to whole faces (Fig. 2). Using the responses to whole faces to localize the regions that distinguish face tokens allowed us to perform further analyses on the responses to the face halves without incurring in bias because, with this procedure, separate datasets are used for localization and for classification within the ROIs. A set of regions was found to encode information that distinguishes between face tokens depicting different people significantly above chance ($P < 0.05$ corrected). These included a right ventral occipital region (peak Montreal Neurological Institute [MNI] coordinates 21 –84 –9), a right ventral fusiform region (peak MNI coordinates 45 –59 –8), and the anterior temporal lobe (ATL) bilaterally (peak MNI coordinates 41 6 –22 and –37 6 –25). Beyond the ventral stream, the right posterior cingulate (peak MNI coordinates 12 –59 14) and the intraparietal sulcus (IPS) bilaterally (peak MNI coordinates 32 –65 34 and –20 –67 37) were also found to encode information that distinguishes between face tokens depicting different people ($P < 0.05$ corrected).

Sensitivity and Invariance of Face Representations to Different Partial Views of Faces

Having individuated regions that encode information that distinguishes between face tokens depicting different people with the responses to the whole-face stimuli, we proceeded to investigate the sensitivity and invariance of face representations across

different partial views of the faces within 12-mm spherical ROIs centered in the information peaks and in an additional right V1 ROI generated anatomically (see Materials and Methods). SVMs were used to perform a 4-way classification between the 4 different halves (upper, lower, right, left) in each of the ROIs, collapsing across different identities (Fig. 3). Since the classification performed is 4 way, chance is at 25%. Within the ventral stream ROIs, significant classification of face halves was found in V1 (accuracy = 34.42%, $t_{(12)} = 4.02$, $P < 0.005$, two tailed) and in the ventral occipital ROI (accuracy = 33.29%, $t_{(12)} = 3.87$, $P < 0.005$, two tailed). A trend was observed in the ventral posterior temporal ROI (accuracy = 27.44%, $t_{(12)} = 2.14$, $P = 0.053$, two tailed), while no significant classification was found in the anterior temporal lobes (right ATL accuracy = 23.60%, $t_{(12)} = -1.14$, $P > 0.1$, two tailed; left ATL accuracy = 24.16%, $t_{(12)} = -0.74$, $P > 0.1$, two tailed). Beyond the ventral stream, classification of face half was nonsignificant in the IPS bilaterally (right IPS accuracy = 25.92%, $t_{(12)} = 0.98$, $P > 0.1$, two tailed; left IPS accuracy = 25.28%, $t_{(12)} = 0.55$, $P > 0.1$, two tailed) but was significant in the right posterior cingulate (accuracy = 28.73%, $t_{(12)} = 4.86$, $P < 0.005$, two tailed).

In order to further investigate the effect of face halves, we attempted to reconstruct the pattern of response in a ROI to the whole face as a linear combination of the patterns of response to 2 halves that constitute that face when they were presented in isolation. For instance, we attempted to reconstruct the pattern of response to the whole face in a ROI as a linear combination of the pattern of response to the left half and the pattern of response to the right half (Fig. 4, see Materials and Methods). If the response to the whole face is identical to the response to the left half and different from that of the right half, the coefficient for the left half will be 1 and that for the right half 0, and vice versa. The greater the contribution of the response to one of the halves, the greater the corresponding coefficient will be.

In posterior regions of the ventral stream, we expected a preference for the face halves presented in the contralateral half of the visual field (as discussed in the Materials and Methods section, the left halves of the faces appeared in the right visual hemifield, while the right halves of the faces appeared in the left visual hemifield, just as they would within the context of a centrally

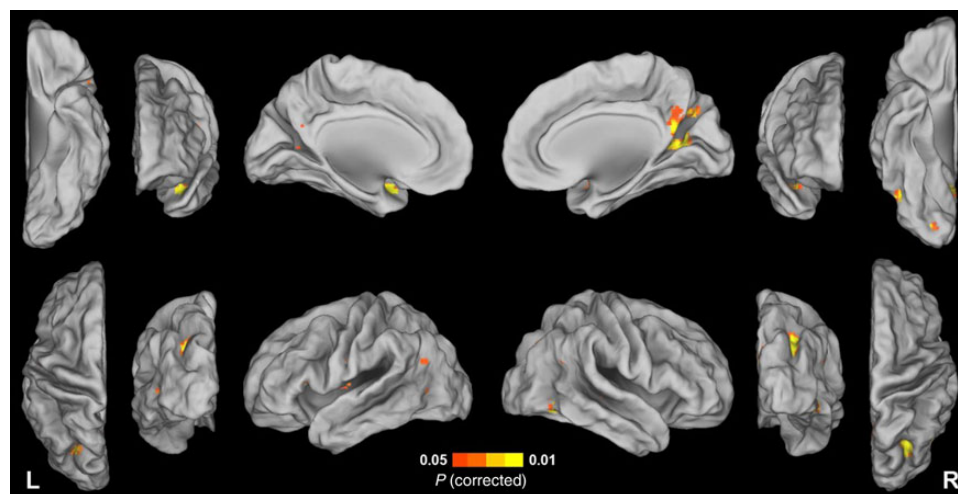


Figure 2. Map of the most informative voxels for classifying between the whole faces of the distractor identities ($P < 0.05$ corrected). Informative voxels were found in a right ventral occipital region (peak Montreal Neurological Institute [MNI] coordinates 21 –84 –9), a right ventral fusiform region (peak MNI coordinates 45 –59 –8), the anterior temporal lobe (ATL) bilaterally (peak MNI coordinates 41 6 –22 and –37 6 –25), the right posterior cingulate (peak MNI coordinates 12 –59 14), and the intraparietal sulcus bilaterally (peak MNI coordinates 32 –65 34 and –20 –67 37).

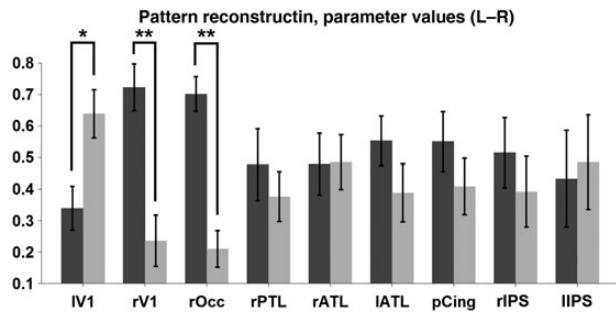


Figure 3. Parameter values for the reconstruction of the response pattern to the whole face as a linear combination of the response patterns to the left and right face halves. The parameter for the left face is shown in dark gray, the parameter for the right face in light gray. A significantly greater contribution of the response pattern to the contralateral half was found in V1 (right V1: $t_{(12)} = 3.27$, one tailed $P < 0.005$; left V1: $t_{(12)} = -2.1415$, one tailed $P < 0.05$) and right ventral occipital cortex ($t_{(12)} = 4.87$, one tailed $P < 0.005$).

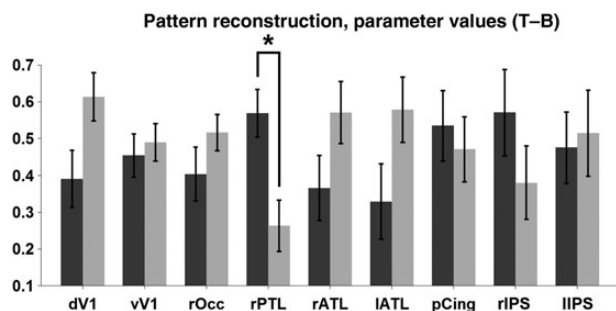


Figure 4. Parameter values for the reconstruction of the response pattern to the whole face as a linear combination of the response patterns to the top and bottom face halves. The parameter for the top face is shown in dark gray, the parameter for the bottom face in light gray. A significantly greater contribution of the response pattern to the top half was found in the right posterior temporal lobe ($t_{(12)} = 2.41$, $P < 0.05$).

presented whole face). In support of this hypothesis, a significant difference between the coefficients for the left and right halves was found, with greater coefficients for the face half presented in the contralateral half of the visual field (right V1: $t_{(12)} = 3.27$, one tailed $P < 0.005$; left V1: $t_{(12)} = -2.1415$, one tailed $P < 0.05$; right ventral occipital: $t_{(12)} = 4.87$, one tailed $P < 0.005$). While ventral stream regions showing significant classification between face halves (V1 and posterior occipital) showed a preference for the contralateral half of the visual field, the right posterior cingulate, which also showed significant classification between different face halves, did not show a preference for the contralateral half of the visual field ($t_{(12)} = 0.78$, one tailed $P > 0.1$). Note that the right posterior cingulate ROI, despite being close to the falx cerebri, was entirely within the right hemisphere. Therefore, the lack of preference for the contralateral half of the visual field could not be due to accidental averaging of right hemisphere voxels showing a left visual hemifield preference and of left hemisphere voxels showing a right visual hemifield preference. No other region showed a significant difference between the coefficients for the face halves presented in the contralateral and in the ipsilateral halves of the visual field (ventral posterior temporal: $t_{(12)} = 0.52$, $P > 0.1$; right ATL: $t_{(12)} = -0.03$, $P > 0.1$; left ATL: $t_{(12)} = 0.99$, $P > 0.1$; right IPS: $t_{(12)} = 0.56$, $P > 0.1$; left IPS: $t_{(12)} = -0.17$, $P > 0.1$, all one tailed). An analogous investigation of the contributions of the top and bottom halves to the responses

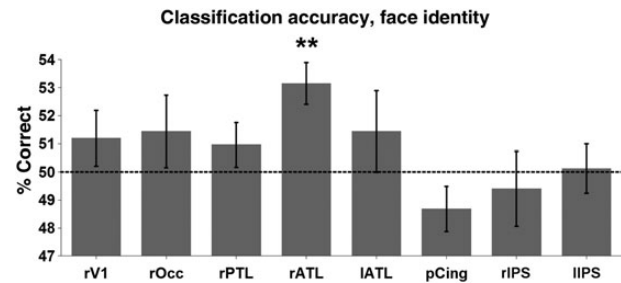


Figure 5. Accuracy of the classification between the 2 distractor identities, generalizing across different face halves. Chance is at 50%. The classifier (SVM) was trained to discriminate between the 2 distractor identities using data from 3 of the halves, and was tested to discriminate between those identities on data from the fourth half. This procedure was iterated for all choices of the test half. Only the right ATL performed significantly above chance (accuracy = 53.13%, $t_{(12)} = 4.16$, $P < 0.005$).

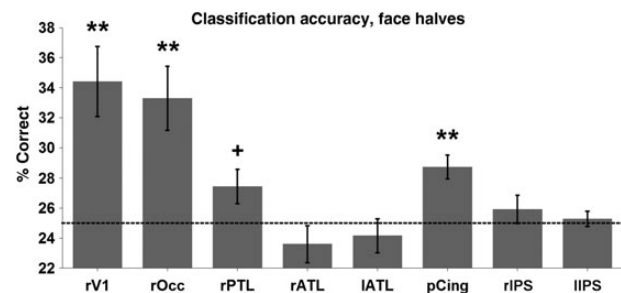


Figure 6. Accuracy of the classification between the 4 halves, collapsing across identity. Chance is at 25%. Accuracy significantly above chance was found in primary visual cortex (accuracy = 34.42%, $t_{(12)} = 4.02$, $P < 0.005$), ventral occipital cortex (accuracy = 33.29%, $t_{(12)} = 3.87$, $P < 0.005$), and posterior cingulate (accuracy = 28.73%, $t_{(12)} = 4.86$, $P < 0.005$). A trend was observed in the posterior temporal lobe (accuracy = 27.44%, $t_{(12)} = 2.14$, $P = 0.053$).

to the whole faces was performed (Fig. 5), reconstructing the response to the whole faces as a linear combination of the responses to the top half and the bottom half. This analysis individuated a significant difference only in the ventral posterior temporal ROI ($t_{(12)} = 2.41$, two tailed $P < 0.05$), in which the coefficient for the top half was greater than the coefficient for the bottom half, indicating that the top half of the faces contributed more to the representation of whole faces than the bottom half in this region.

The invariance of face representations was investigated by training SVMs to discriminate between the 2 distractor identities based on the responses to 3 of the 4 halves, and testing the ability to generalize the identity classification to the fourth remaining half (Fig. 6). This procedure allows the use of the responses to different stimuli for training and for testing of the classifiers, and was iterated for all choices of the face half that was left out for testing. Identity classification with invariance across changes in a face half was found in the right ATL (accuracy = 53.13%, $t_{(12)} = 4.16$, $P < 0.005$, two tailed). No other region showed identity classification with invariance across changes in a face half (right V1: accuracy = 51.20%, $t_{(12)} = 1.22$, $P > 0.1$; right ventral occipital: accuracy = 51.44%, $t_{(12)} = 1.12$, $P > 0.1$; right ventral posterior temporal: accuracy = 50.96%, $t_{(12)} = 1.20$, $P > 0.1$; left ATL: accuracy = 51.44%, $t_{(12)} = 1.00$, $P > 0.1$; right posterior cingulate: accuracy = 48.68%, $t_{(12)} = -1.64$, $P > 0.1$; right IPS: accuracy = 49.40%, $t_{(12)} = -0.45$, $P > 0.1$; left IPS: accuracy = 50.12%, $t_{(12)} = 0.14$, $P > 0.1$, all two tailed).

Generalization accuracy in right ATL was investigated in greater depth, analyzing separately the accuracy when different halves were used as the target (left half: mean = 0.5144, SEM = 0.0142; right half: mean = 0.5288, SEM = 0.0160; bottom half: mean = 0.5433, SEM = 0.0174; top half: mean = 0.5385, SEM = 0.0307). The results broken down by type of face half suggest that classification accuracy was higher when the training set contained more informative face parts. In fact, the face halves are particularly informative (given the symmetry of faces), and when they were both included in the training set classification was highest. Furthermore, in line with this observation, classification was higher when the top half, which contains the eyes, was included in the training set, than when it was in the testing set. However, caution should be exercised in the interpretation of this pattern of results, because the differences between generalization accuracy for different target halves are very subtle.

Functional Connectivity Between Regions Encoding Information That Distinguishes Between Face Tokens Depicting Different People

The investigation of the sensitivity and invariance of face representations was complemented with an analysis of functional connectivity, to gain insights about the relations and interactions between the regions encoding information that distinguish between face tokens depicting different people. Significant functional connectivity was found between the ventral occipital region and the ventral posterior temporal region ($r = 0.57$, $t_{(12)} = 9.36$, corrected $P < 0.005$, two tailed), and between the right and left ATL ($r = 0.50$, $t_{(12)} = 5.57$, corrected $P < 0.005$, two tailed). Strong functional connectivity was found between the parietal regions (right and left IPS: $r = 0.86$, $t_{(12)} = 17.18$, corrected $P < 0.005$; right IPS and posterior cingulate: $r = 0.62$, $t_{(12)} = 10.50$, corrected $P < 0.005$; left IPS and posterior cingulate: $r = 0.68$, $t_{(12)} = 10.98$, corrected $P < 0.005$, all two tailed). Functional connectivity between the ventral posterior temporal region and the bilateral ATL was not as strong, but was highly significant (with right ATL: $r = 0.27$, $t_{(12)} = 4.41$, corrected $P < 0.05$; with left ATL: $r = 0.35$, $t_{(12)} = 5.97$, corrected $P < 0.005$, all two tailed). Functional connectivity between the ventral occipital region and both the right and left ATL, instead, was nonsignificant (with right ATL: $r = 0.19$, $t_{(12)} = 2.73$, corrected $P > 0.1$; with left ATL: $r = 0.24$, $t_{(12)} = 3.11$, corrected $P > 0.1$, all two tailed). A particularly interesting aspect of the functional connectivity results was the high connectivity observed between posterior ventral stream regions (ventral occipital and ventral posterior temporal) and parietal regions (ventral occipital and posterior cingulate: $r = 0.40$, $t_{(12)} = 5.59$, corrected $P < 0.005$; ventral occipital and right IPS: $r = 0.44$, $t_{(12)} = 5.27$, corrected $P < 0.005$; ventral occipital and left IPS: $r = 0.42$, $t_{(12)} = 4.60$, corrected $P < 0.05$; ventral posterior temporal and posterior cingulate: $r = 0.53$, $t_{(12)} = 12.27$, corrected $P < 0.005$; ventral posterior temporal and right IPS: $r = 0.71$, $t_{(12)} = 17.47$, corrected $P < 0.005$; ventral posterior temporal and left IPS: $r = 0.64$, $t_{(12)} = 13.75$, corrected $P < 0.005$, all two tailed). Functional connectivity between the bilateral ATL and the posterior cingulate was significant (right ATL and posterior cingulate: $r = 0.30$, $t_{(12)} = 4.44$, corrected $P < 0.05$; left ATL and posterior cingulate: $r = 0.39$, $t_{(12)} = 4.70$, corrected $P < 0.05$, all two tailed), but functional connectivity between the bilateral ATL and the bilateral IPS was nonsignificant (right ATL and right IPS: $r = 0.20$, $t_{(12)} = 2.60$, corrected $P > 0.1$; right ATL and left IPS: $r = 0.22$, $t_{(12)} = 3.05$, corrected $P > 0.1$; left ATL and right IPS: $r = 0.27$, $t_{(12)} = 3.11$, corrected $P > 0.1$; left ATL and left IPS: $r = 0.32$, $t_{(12)} = 3.52$, corrected $P = 0.09$, all two tailed).

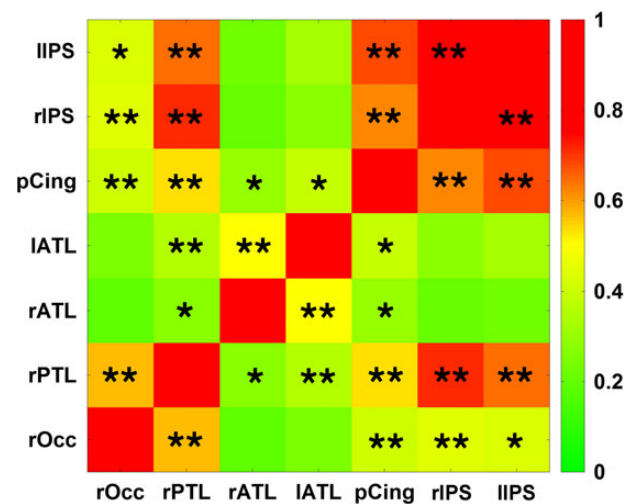


Figure 7. Functional connectivity between regions containing the most informative voxels for classifying between the whole faces of the distractor identities. Significant functional connectivity is marked with * $P < 0.05$ and with ** $P < 0.005$.

Discussion

In this study, we found a network of regions that encode information that distinguishes between face tokens depicting different people (Fig. 2). The network includes ventral stream regions previously reported to encode information that distinguishes between face tokens depicting different people (Nestor et al. 2011, 2014; Anzellotti et al. 2013; Goesaert and de Beeck 2013; Verosky et al. 2013; Cowen et al. 2014). In addition, the network includes regions beyond the ventral stream (posterior cingulate and IPS), some of which are known to be involved in face processing (Haxby et al. 2000; Gobbini and Haxby 2006), but were not known to encode information that distinguishes between face tokens depicting different people. This is because most MVPAs to investigate representations that distinguish between face tokens depicting different people were restricted to the ventral visual stream (Natu et al. 2010; Nestor et al. 2011; Anzellotti et al. 2013). Posterior cingulate has been previously implicated in semantic knowledge and person concepts (Fairhall and Caramazza 2013a, 2013b), theory of mind (Saxe and Powell 2006; Mitchell 2008), and episodic memory (Wagner et al. 2005); IPS has been implicated in short-term memory for objects (Xu and Chun 2005; Xu 2007) and in object identification (Xu 2009; Jeong and Xu 2013).

Of primary importance in the present study is that face representations invariant across changes in face halves were found in the right ATL (Fig. 6). In a complementary analysis, information about identity-irrelevant properties in the ventral stream was found to decline along the posterior–anterior axis (Fig. 3). Functional connectivity revealed 3 clusters of highly connected regions: 1) occipitotemporal regions, 2) the left and right ATL, and 3) parietal regions (Fig. 7).

Invariance of Face Representations

The right ATL was the only region found to encode invariant face representations (Fig. 6). This finding is consistent with previous fMRI research on face representations (Nestor et al. 2011; Anzellotti et al. 2013) and with results in monkey physiology showing that the ATL is the region in the ventral stream where face representations achieve maximal invariance to identity-irrelevant

properties (Freiwald and Tsao 2010). Furthermore, it is consistent with patient studies showing that damage to the right anterior temporal lobe is associated with deficits for the recognition of faces (Tranel et al. 1997). The right ATL is not involved exclusively in computing invariance across different face parts. In a previous study (Anzellotti et al. 2013), we found evidence of face representations invariant across changes in viewpoint in the right ATL and occipitotemporal regions. These results indicate that the right ATL is the most convincing candidate region to represent face identity (for an in-depth review of the role of ATL for face recognition see Collins and Olson 2014).

If the right ATL encodes face representations that are invariant to identity-irrelevant properties, why is classification not 100% accurate? It could be hypothesized that this is due to the additional presence within the right ATL of neurons that are sensitive to identity-irrelevant properties (see for instance DiCarlo and Maunsell 2003), whose responses might get averaged with those of neurons showing invariance because of the relatively low spatial resolution of fMRI. A recent study (Issa et al. 2013) has shown that fMRI maps are highly correlated with maps derived from smoothed electrophysiology data (supporting the use of MVPA to investigate the information encoded in a brain region), but that electrophysiology data additionally reveal reliable information encoded at smaller spatial scales, which cannot be extracted from fMRI data. Another possibility is that invariance emerges over time, and that the low temporal resolution of fMRI averages late invariant responses with earlier responses that are more sensitive to identity-irrelevant properties. A combination of these accounts is also possible, in which the activation of a distinct neural population within the right ATL that is sensitive to identity-irrelevant properties gets inhibited over time. Accounts of this type are compatible with the finding that the amount of invariant face identity information in the ATL of monkeys varies as a function of time, reaching a peak approximately 300 ms after stimulus onset (Freiwald and Tsao 2010, Figure 4I in that article). However, these accounts do not provide a complete explanation of the size of classification accuracy effects in the fMRI literature. In fact, classification of specific face tokens depicting different people was found to have higher accuracy than invariant classification, but nonetheless remained below 60% for carefully controlled stimuli (Anzellotti et al. 2013). Therefore, the size of classification accuracy effects for invariant classification of face identity is likely to depend on multiple factors.

Training and testing the classifier with entirely nonoverlapping face halves (e.g., training with the top half alone and testing on the bottom half) did not yield significant classification. This might suggest either a partial reliance of classification on visual details of the image, or the lack of adequate power in this analysis given that only one-third of the data were available for training of the classifier (data from one half as opposed to data from 3 halves).

No evidence of classification of identity with invariance across face half was found in the left ATL. This result is in agreement with previous research, which found face representations invariant to differences in viewpoint in the right but not in the left ATL (Anzellotti et al. 2013). Damage to the left ATL is associated with deficits for person naming (Damasio et al. 1996). Therefore, the presence of invariant face representations in the right ATL but not in the left could be due to a division of labor between the right and left ATL in which the right plays a more important role for the representation of faces while the left plays a more important role for the representation of names.

Recent results show that face representations in occipital cortex encode information about face parts, regardless of their

location in the face (Liu et al. 2010). Given these results, one might expect to find classification of identity generalizing across halves in our occipital ROI, since the parts in the images used for testing the classifier were also contained in the training images. However, responses to different face parts in macaque's occipital face patch sum linearly only in early stages of the response (0–40 ms) (Issa and DiCarlo 2012). The BOLD signal depends on neural activity in both early and later stages; therefore, it is expected that the BOLD response to a whole is not just a linear sum of its parts. Furthermore, responses in the macaque's occipital face patch are strongly dependent on the position of the parts in the visual field: face parts in the contralateral visual field were found to drive stronger responses (Issa and DiCarlo 2012). This observation is consistent with the results of the pattern reconstruction analysis we applied to the ventral occipital ROI, which found a greater coefficient associated with the contralateral half of a face. The response to the contralateral half of a face was associated with a greater coefficient also in V1 of each hemisphere, as expected because of its retinotopic organization.

Sensitivity of Face Representations to Identity-Irrelevant Properties

We complemented the analysis of information about invariant face identity with an analysis of information about the different partial views of the faces. Sensitivity to differences in the partial views of the faces was found in posterior ventral stream regions (V1 and ventral occipital) and in the posterior cingulate. In the posterior ventral stream regions, a hemifield bias was evident, with regions showing a preference for face halves presented in the contralateral half of the visual field. This finding is consistent with earlier reports of a preference for contralateral stimuli in multiple regions in the ventral stream, including the occipital face area and the fusiform face area (Hemond et al. 2007). Interestingly, no evidence of preference for stimuli presented in the contralateral visual field was found in the ATL. This suggests that information about object location in the visual field might no longer be represented explicitly in the ATL. However, caution should always be exercised in the interpretation of negative results.

In addition to posterior ventral stream regions, the posterior cingulate was found to encode information about the different halves. The posterior cingulate has been implicated in episodic memory and in social cognition (Fletcher et al. 1995; Wagner et al. 2005; Saxe and Powell 2006; Mitchell 2008). A recent study (Fairhall et al. 2013) has shown that it encodes information about the domain (e.g., people, places) to which individual objects belong. Furthermore, information about the domain is present for both picture and word stimuli (Fairhall and Caramazza 2013a, 2013b). In addition, the posterior cingulate shows differential functional connectivity with the medial ATL during memory tasks (O'Neil et al. 2012). Given these results, it seems unlikely that the posterior cingulate differentiates between different face halves on the basis of low-level visual information. One possible hypothesis is that the posterior cingulate encodes information about the details of the image perceived in order to contribute to the subsequent retrieval of detailed episodic information. An alternative possibility is that distinct patterns of response to different face halves in the posterior cingulate are due to the posterior cingulate's sensitivity to the social relevance of different face parts (e.g., eyes, mouth) present in different halves.

The absence of occluders permitted an investigation of the response to a whole face as a combination of the responses to face

halves. However, the use of such stimuli might have prompted participants to adopt task-specific strategies for identity recognition, such as engaging in face completion. If participants had adopted such a strategy, resulting in strong face completion effects, classification of identity across different face halves would have been expected also in other face-selective regions besides the right ATL, and potentially even in V1. Instead, classification of face identity across different face halves was only observed in the right ATL. Future studies should employ occluders to fully rule out any contribution of face completion effects.

Functional Connectivity

Functional connectivity analyses revealed that the right ATL, which encodes information about face identity generalizing across face halves, shows particularly strong connectivity with the left ATL. Damage to the left ATL has been associated with deficits for person naming (Damasio et al. 1996). Connectivity between the right and left ATL might reflect communication between the 2 regions that contributes to the retrieval of a person's name from her face, and of the person's face from her name.

Conclusion

In this study, the right ATL was found to encode information about face identity generalizing across face halves. Together with previous findings showing that the right ATL encodes information about face identity generalizing across changes in viewpoint (Anzellotti et al. 2013), these results indicate the right ATL as the most convincing candidate region for the representation of face identity (see Anzellotti and Caramazza 2014). These results do not exclude the possibility that right ATL might encode other information about person identity, such as semantic knowledge. High functional connectivity was observed between the right ATL and the left ATL, which has been implicated in the retrieval of people's names (Damasio et al. 1996). Connectivity between right and left ATL might underlie the ability to retrieve a person's name given their face and vice versa. Consistent with a hierarchical view of face processing, low-level, identity-irrelevant information was found to decline along the posterior–anterior axis in the ventral stream.

Funding

This work was supported by the Provincia Autonoma di Trento and by the Fondazione Cassa di Risparmio di Trento e Rovereto. S.A. was supported by a grant from the Fondazione Cassa Rurale di Trento.

Notes

We thank Valentina Brentari for assistance in collecting data. *Conflict of Interest:* None declared.

References

- Anzellotti S, Caramazza A. 2014. The neural mechanisms for the recognition of face identity in humans. *Front Psychol.* 5:672 (Last accessed 19 January 2015).
- Anzellotti S, Fairhall SL, Caramazza A. 2013. Decoding representations of face identity that are tolerant to rotation. *Cereb Cortex.* 24:1988–1995.
- Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage.* 16:S497.
- Chang CC, Lin CJ. 2011. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol.* 2:27:1–27:27.
- Collins JA, Olson IR. 2014. Beyond the FFA: the role of the ventral anterior temporal lobes in face processing. *Neuropsychologia.* 61:65–79.
- Cowen AS, Chun MM, Kuhl BA. 2014. Neural portraits of perception: reconstructing face images from evoked brain activity. *Neuroimage.* 94:12–22.
- Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR. 1996. A neural basis for lexical retrieval. *Nature.* 380(6574):499–505.
- Declaration of Helsinki. 1996. *BMJ.* 335:624.
- De Martino F, Valente G, Staeren NL, Ashburner J, Goebel R, Formisano E. 2008. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage.* 43:44.
- DiCarlo JJ, Maunsell JH. 2003. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol.* 89:3264–3278.
- Fairhall SL, Anzellotti S, Ubaldi S, Caramazza A. 2013. Person- and place-selective neural substrates for entity-specific semantic access. *Cereb Cortex.* 24:1687–1696.
- Fairhall SL, Caramazza A. 2013a. Brain regions that represent amodal conceptual knowledge. *J Neurosci.* 33:10552–10558.
- Fairhall SL, Caramazza A. 2013b. Category-selective neural substrates for person- and place-related concepts. *Cortex.* 49:2748–2757.
- Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD. 1995. Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition.* 57:109–128.
- Freiwald WA, Tsao DY. 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science.* 330:845–851.
- Gobbini MI, Haxby JV. 2006. Neural response to the visual familiarity of faces. *Brain Res Bull.* 71:76–82.
- Goesaert E, de Beeck HPO. 2013. Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. *J Neurosci.* 33:8549–8558.
- Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn Sci.* 4:223–233.
- Hemond CC, Kanwisher NG, de Beeck HPO. 2007. A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS ONE.* 2:e574.
- Issa EB, DiCarlo JJ. 2012. Precedence of the eye region in neural processing of faces. *J Neurosci.* 32:16666–16682.
- Issa EB, Papanastassiou AM, DiCarlo JJ. 2013. Large-scale, high-resolution neurophysiological maps underlying fMRI of macaque temporal lobe. *J Neurosci.* 33:15207–15219.
- Jeong SK, Xu Y. 2013. Neural representation of targets and distractors during object individuation and identification. *J Cogn Neurosci.* 25:117–126.
- Kietzmann TC, Swisher JD, König P, Tong F. 2012. Prevalence of selectivity for mirror-symmetric views of faces in the ventral and dorsal visual pathways. *J Neurosci.* 32:11763–11772.
- Liu J, Harris A, Kanwisher N. 2010. Perception of face parts and face configurations: an fMRI study. *J Cogn Neurosci.* 22:203–211.
- Logothetis NK, Pauls J. 1995. Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cereb Cortex.* 5(3):270–288.
- MacEvoy SP, Epstein RA. 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol.* 19:943–947.
- Mitchell JP. 2008. Contributions of functional neuroimaging to the study of social cognition. *Curr Dir Psychol Sci.* 17:142–146.

- Natu VS, Jiang F, Narvekar A, Keshvari S, Blanz V, O'Toole AJ. 2010. Dissociable neural patterns of facial identity across changes in viewpoint. *J Cogn Neurosci*. 22:1570–1582.
- Nestor A, Plaut D, Behrmann M. 2014. Facial identity an investigation of neural encoding and image reconstruction. *J Vis*. 14: article 604.
- Nestor A, Plaut DC, Behrmann M. 2011. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc Natl Acad Sci USA*. 108:9998–10003.
- O'Neil EB, Protzner AB, McCormick C, McLean DA, Poppenk J, Cate AD, Köhler S. 2012. Distinct patterns of functional and effective connectivity between perirhinal cortex and other cortical regions in recognition memory and perceptual discrimination. *Cereb Cortex*. 22:74–85.
- Rorden C, Brett M. 2000. Stereotaxic display of brain lesions. *Behav Neurol*. 12:191–200.
- Sadr J, Jarudi I, Sinha P. 2003. The role of eyebrows in face recognition. *Perception*. 32:285–293.
- Saxe R, Powell LJ. 2006. It's the thought that counts specific brain regions for one component of theory of mind. *Psychol Sci*. 17:692–699.
- Tanaka JW, Farah MJ. 1993. Parts and wholes in face recognition. *Q J Exp Psychol*. 46:225–245.
- Tranel D, Damasio H, Damasio AR. 1997. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*. 35: 1319–1327.
- Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. 2010. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J Neurophysiol*. 103:297–321.
- Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH. 2001. An integrated software suite for surface-based analyses of cerebral cortex. *J Am Med Inform Assoc*. 8:443–459.
- Verosky SC, Todorov A, Turk-Browne NB. 2013. Representations of individuals in ventral temporal cortex defined by faces and biographies. *Neuropsychologia*. 51:2100–2108.
- Wagner AD, Shannon BJ, Kahn I, Buckner RL. 2005. Parietal lobe contributions to episodic memory retrieval. *Trends Cogn Sci*. 9:445–453.
- Xu Y. 2009. Distinctive neural mechanisms supporting visual object individuation and identification. *J Cogn Neurosci*. 21:511–518.
- Xu Y. 2007. The role of the superior intraparietal sulcus in supporting visual short-term memory for multifeature objects. *J Neurosci*. 27:11676–11686.
- Xu Y, Chun MM. 2005. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*. 440: 91–95.
- Young AW, Hellawell D, Hay DC. 1987. Configurational information in face perception. *Perception*. 16:747–759.
- Zaitsev M, Hennig J, Speck O. 2004. Point spread function mapping with parallel imaging techniques and high acceleration factors: fast, robust, and flexible method for echo-planar imaging distortion correction. *Magn Reson Med*. 52:1156–1166.