

Decoding Representations of Face Identity That are Tolerant to Rotation

Stefano Anzellotti^{1,2}, Scott L. Fairhall² and Alfonso Caramazza^{1,2}

¹Department of Psychology, Harvard University, Cambridge, MA 02138, USA and ²Center for Mind/Brain Sciences (CIMeC), University of Trento, 38123 Trento, Italy

Address correspondence to Stefano Anzellotti, Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA. Email: anzellot@fas.harvard.edu

In order to recognize the identity of a face we need to distinguish very similar images (specificity) while also generalizing identity information across image transformations such as changes in orientation (tolerance). Recent studies investigated the representation of individual faces in the brain, but it remains unclear whether the human brain regions that were found encode representations of individual images (specificity) or face identity (specificity plus tolerance). In the present article, we use multivoxel pattern analysis in the human ventral stream to investigate the representation of face identity across rotations in depth, a kind of transformation in which no point in the face image remains unchanged. The results reveal representations of face identity that are tolerant to rotations in depth in occipitotemporal cortex and in anterior temporal cortex, even when the similarity between mirror symmetrical views cannot be used to achieve tolerance. Converging evidence from different analysis techniques shows that the right anterior temporal lobe encodes a comparable amount of identity information to occipitotemporal regions, but this information is encoded over a smaller extent of cortex.

Keywords: conspecifics, fMRI, individual, invariance, viewpoint

Introduction

In our daily life, we constantly need to recognize objects in order to act appropriately in the world. In some cases, it is sufficient to recognize objects as instances of a particular type (e.g., a hammer, a dog), but sometimes we need to recognize specific individuals (e.g., the friend we set out to meet, our home). The recognition of individuals poses significant computational challenges, because it requires detection of very subtle differences (specificity) while at the same time being tolerant for large variations in sensory stimulation that naturally occur in the world like rotations, translations and scalings. The category of objects that we need to recognize at the individual level most frequently is probably people and, in order to recognize people, visual information about their faces is particularly important. This article investigates the specificity and tolerance of face representations in the ventral visual stream.

Neuropsychological studies of face recognition (Meadows 1974; Damasio et al. 1996; Tranel et al. 1997) have reported the existence of patients with selective difficulties in the recognition and/or naming of individuals following damage to occipitotemporal cortex and to the anterior portions of the temporal lobe (ATL). Single-cell recording studies in humans (Quiroga et al. 2005) have found neurons responding to a person's identity with tolerance across image changes in the hippocampus, which receives afferents from the ATL via the entorhinal cortex (Lopes da Silva and Arnolds 1978).

Neuroimaging studies of face processing have reported stronger responses to faces than to other categories of objects in the fusiform face area (FFA, Sergent et al. 1992; Puce et al. 1995, 1996; Kanwisher et al. 1997), the occipital face area (OFA; Kanwisher et al. 1997; Gauthier et al. 2000), and the right ATL (Rajimehr et al. 2009). It remains an open question whether these regions encode specific and tolerant representations of individual faces.

Two recent studies used multivoxel pattern analysis (MVPA) to investigate the specificity of face representations. In 1 study, Kriegeskorte et al. (2007) presented images of 2 faces in three-fourth orientation, and found that the right ATL contains information that can distinguish between 2 face images (specificity). However, the results of that experiment do not allow conclusions regarding transformation-tolerant representations because only 1 image for each identity was used; therefore, there were no transformations for which tolerance could be tested. In another study, Nestor et al. (2011), after averaging together the blood oxygen level-dependent (BOLD) responses corresponding to a same face with different expressions, identified a set of ventral stream regions including the right ATL and the fusiform gyrus bilaterally that contain information about individual images of faces. For each identity, multiple face images with different expressions were used. However, this second study, too, allows only limited conclusions since changes in expression only affect some parts of a face, and classification could rely on the parts of the faces that remain unchanged. Furthermore, for each identity, the responses to the same stimuli were used for training and testing the classifiers; therefore, the results might reflect representations of specific images, rather than the transformation-tolerant representations needed for recognition of individuals.

These studies constitute an important first step in the investigation of face representations in the ventral stream. However, they did not test the tolerance of face representations, which is crucial for our interaction with the world. Another aspect of face recognition that has not been thoroughly tested is the role of mirror symmetry for generalization. It has been proposed that mirror symmetrical images might play an important role for generalization across viewpoint in object recognition (Vetter et al. 1994). The existence of neurons selective for mirror symmetrical views of objects has been reported in ventral temporal cortex (Logothetis et al. 1995; Freiwald and Tsao 2010). However, it remains an open question whether generalization in ventral temporal cortex is limited uniquely to mirror reflections, or whether ventral temporal cortex instead encodes representations that generalize across changes in orientation other than mirror reflections.

The tolerance of face representations has been investigated in depth in monkeys. In a recent study, Freiwald and Tsao (2010) used fMRI to localize regions in the monkey brain

showing stronger BOLD responses to faces than to other objects, and then targeted these regions for electrophysiological analysis. They record neural responses to faces of different identities seen from different viewpoints, and report finding cells that exhibit increasing tolerance to changes in viewpoint in moving from posterior to anterior areas in the ventral stream, with the highest tolerance achieved in a region denominated anterior medial (AM). In the present study, using fMRI and MVPA, we investigate where tolerance across viewpoints is achieved in the human ventral stream. For the MVPA analysis, we used 2 different approaches: a region of interest (ROI) approach and a feature selection-based approach, because each of the approaches has important advantages. The ROI approach makes it easier to compare the results obtained in our study with those obtained in previous studies. However, it cannot tell us whether information is only present within our ROIs, or whether it is also present outside the ROIs. The feature selection-based approach gives us this additional piece of information.

Materials and Methods

Participants

Ten participants (age range 18–50 years, mean 27.1 years) took part in the experiment. The participants' consent was obtained according to the Declaration of Helsinki. The project was approved by the Human Subjects Committees at the University of Trento and Harvard University. Data from one participant were discarded from the analysis because of poor performance during a behavioral training session administered on the day before the scanning.

Stimuli

Images of 5 face identities at 5 different orientations were generated rendering 3D models faces in DAZ-3D (Fig. 1A, Supplementary Fig. 1). The use of 3D models ensured that information about rotation angle, color, and texture could not be used to distinguish between the faces. Stimuli were presented with Psychtoolbox (Brainard 1997; Pelli 1997) running on MATLAB, with the add-on ASF (Schwarzbach 2011), using an Epson EMP 9000 projector. Images were projected on

a frosted screen at the top of the bore, viewed through a mirror attached to the head coil.

Experimental Design

Before entering the scanner, participants were briefly familiarized with the 5 faces seen from 5 orientations. One face identity (constant across participants) was designated as the “target,” and participants were instructed to respond with the index finger of the right hand to the target face and with the middle finger to the other “distractor” faces (Supplementary Fig. 1). All analyses were performed on the distractor faces; therefore, classification of different distractor faces cannot be attributed to the production of different motor responses. Each trial consisted of the presentation of a face image (500 ms) followed by a fixation cross (1500 ms). The experiment was composed of 3 12-min runs, each composed of ~320 trials. The order of presentation of the stimuli was optimized for deconvolution with optseq2 (<http://surfer.nmr.mgh.harvard.edu/optseq/>). On the day before the scanning, participants took part in a training session (~30 min) during which they were shown rotation videos of 2 of the distractor faces and performed a 1-back identity discrimination task. We did not observe significant training-dependent differences in generalization; therefore, analyses for trained and untrained faces are collapsed together. To avoid any biases in classification, only classification between faces with the same level of training was performed; therefore, differences in training cannot explain the observed classification performance. A block-design functional localizer with faces, houses, and scrambled images was administered at the beginning of the fMRI session. None of the faces shown in the localizer were presented during the other parts of the experiment.

Data Acquisition and Analysis

MRI Scanning Parameters

The data were collected on a Bruker BioSpin MedSpec 4T at the Center for Mind/Brain Sciences (CIMEC) of the University of Trento using a USA Instruments 8-channel phased-array head coil. Before collecting functional data, a high-resolution ($1 \times 1 \times 1 \text{ mm}^3$) T_1 -weighted MPRAGE 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).

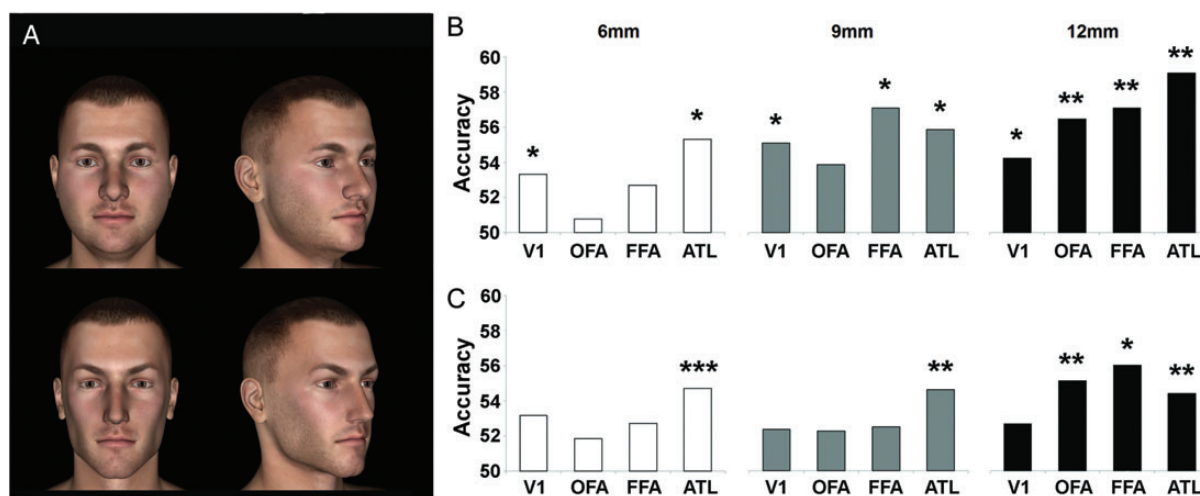


Figure 1. (A) Four example stimuli used for classification, generated with 3D graphics software. All stimuli are shown in Supplementary Figure 1. (B) Specific classification accuracy as a function of ROI size. (C) Tolerant classification accuracy as a function of ROI size. Tolerant classification was performed employing a cross-validation procedure in which a pattern classifier was trained to discriminate 2 face identities with data from 4 of the 5 viewpoints, and tested on the discrimination between the 2 identities seen in the fifth viewpoint, which was not used for training. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$.

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 3×3 mm in plane resolution). Over 3 runs, 1095 volumes of 43 slices were acquired in the axial plane aligned along the long axis of the temporal lobe.

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 2010a, and with custom MATLAB software using the MATLAB bioinformatics toolbox.

Preprocessing. The first 4 volumes of each run were discarded and all images were corrected for head movement. Slice-acquisition delays were corrected using the middle slice as reference. Images were normalized to the standard SPM8 EPI template and resampled to a 3-mm isotropic voxel size. The BOLD signal was high pass filtered at 128 s and prewhitened using an autoregressive model AR(1).

General linear model. Data were modelled with one regressor for the target face and separate regressors for each combination of identity and orientation for distractor faces. Subsequent repetitions (within a run) of the distractor faces were modelled in groups of 3 to balance quality of fit and accurate deconvolution. Regressors were convolved with a standard hemodynamic response function. A parametric modulator for reaction time and 6 motion regressors were included in the model.

Regions of interest definition. ROIs for the OFA, FFA, and ATL were defined with an independent functional localizer individuating the peaks showing stronger activity for faces than for houses. After determining the location of the peaks, spheres of 6, 9, and 12 mm radius were generated centered in each of the peaks, to investigate the effect of radius size and number of voxels considered on classification accuracy. As a control, V1 ROIs were defined using the MARSBAR AAL atlas. Three V1 ROIs were generated, matched in number of voxels to the 6, 9, and 12 mm spherical ROIs. Given that V1 is stripe-shaped, these ROIs were generated starting from the foveal (posterior) portion of V1, and including more voxels progressing peripherally (anteriorly) within the shape constraints determined by the MARSBAR calcarine ROIs (Supplementary Fig. 2). The ROIs for different regions do not overlap at any sphere size. Hippocampus ROIs were defined using the Wake Forest University PickAtlas.

Training and testing of the classifiers. In the first ROI analysis (Fig. 1B), data were divided in 2 independent sets (both comprising all orientations), one used for training and the other for testing. In all other analyses with the exception of those without mirror symmetrical views (Fig. 4), classifiers were trained to discriminate 2 face identities with all data from 4 of the 5 orientations shown, and tested with the orientation not used for training. In the analyses without mirror symmetrical views, one orientation and its mirror symmetrical view were chosen for testing, and the other orientations were used for training. Linear SVM were used for all classifications, as implemented in the MATLAB functions “svmtrain” and “svmclassify.” In all analyses, pairwise discriminations were performed; therefore, chance is at 50%.

Ventral stream recursive feature elimination analysis. Two anatomical ROIs, one for the left ventral stream and one for the right, were generated using the Wake Forest University PickAtlas toolbox for SPM (<http://fmri.wfubmc.edu/software/PickAtlas>). The ROIs extended from the occipital lobe to the anterior temporal pole, and extended medially to include the fusiform gyrus. A 2-stage feature selection was applied (De Martino et al. 2008), with a first stage of mass-univariate selection preserving 1000 voxels, followed by a recursive feature elimination (RFE) analysis (Guyon et al. 2002), that eliminated an additional 800 voxels. A pattern of fMRI activity consists of the BOLD

signal in a number of different voxels, and can be considered as a vector V in a space with as many dimensions as the number of voxels, where each dimension corresponds to one of the voxels. In each voxel, different stimuli will elicit different amounts of BOLD signal: the BOLD signal elicited in a voxel X (that corresponds to dimension D) by a stimulus S is the value of the component of vector V along dimension D . In RFE, a linear SVM is trained on a subset of the data, and the training procedure leads to the individuation of a separating hyperplane that leaves on one side of the space the data points belonging to one class and on the other side of the space the data points belonging to the other class. The way in which the separating hyperplane is tilted in the space (and consequently which data points lie on one side or the other of the hyperplane) depends on the hyperplane's normal vector (and vice versa). This vector has one component along each dimension of the space. For each voxel, the absolute value of the component of the hyperplane's normal vector along that voxel reflects how much the BOLD signal in that voxel contributes to classification. Eliminating the BOLD signal in that voxel from the patterns is equivalent to setting the component of the normal vector along that voxel to zero. The greater the absolute value of the component of the normal vector along that voxel, the greater the change in the direction of the normal vector when that component is set to zero. Therefore, eliminating from the patterns voxels along which the component of the hyperplane's normal vector is smaller will produce a smaller change in the tilt of the hyperplane, and will be less likely to change which data points lie on one side of the hyperplane or the other. By contrast, eliminating voxels along which the component of the hyperplane's normal vector is large will produce large changes in the tilt of the hyperplane, which can lead to large changes in the hyperplane's tilt. This in turn can lead to large drops in classification accuracy, because the original tilt of the hyperplane was selected to separate “at best” (see Boser et al. 1992) the 2 classes of data points. On the basis of these considerations, RFE uses the absolute value of the components of the separating hyperplane's normal vector along the different voxels to evaluate the importance of the contribution of each voxel to classification. At each step, the voxel with the smallest absolute value was eliminated from the patterns, and the procedure (including retraining of the classifier to evaluate the new separating hyperplane) was iterated until the desired number of voxels was reached. The total number of voxels to be kept (400, 200 in each hemisphere) was chosen to match previous work (e.g., Nestor et al. 2011) in order to facilitate comparison. In order to ensure that the test data played no role in the selection of the 200 voxels in each hemisphere (to avoid biases in the analysis), feature selection was repeated for each cross-validation iteration using exclusively the training data.

Probability of detecting discriminative voxels. For each participant and hemisphere, a map of the probability of each voxel to be among the 200 most informative voxels selected by the feature selection procedure was computed. The ventral stream ROIs were subdivided into 14 bins each spanning 3 voxels along the posterior to anterior dimension, covering a range of MNI coordinates between $y = -100$ and $y = 26$. For each bin the sum of the probability values for voxels in that bin was computed and normalized by the number of voxels in the bin to control for differences in the size of the mask at different levels along the posterior to anterior axis. The mean and the standard error of the mean of the normalized probabilities for the 9 participants were then computed.

Clusters definition. Anterior and posterior clusters in the ventral stream were defined on the basis of the plots of the probability of finding discriminative voxels (Fig. 2B). The plots show 2 groups of contiguous bins in each hemisphere showing above-chance probability of finding informative voxels, between MNI coordinates -73 and -46 , and between MNI coordinates 8 and 17. For each group of contiguous bins showing above-chance probability of finding informative voxels, we created a ROI defined as the set of the most informative voxels found with RFE in that group of bins. The selection of the voxels to be included in the ROIs was based on data in the training set, thus avoiding the risk of “double dipping.”

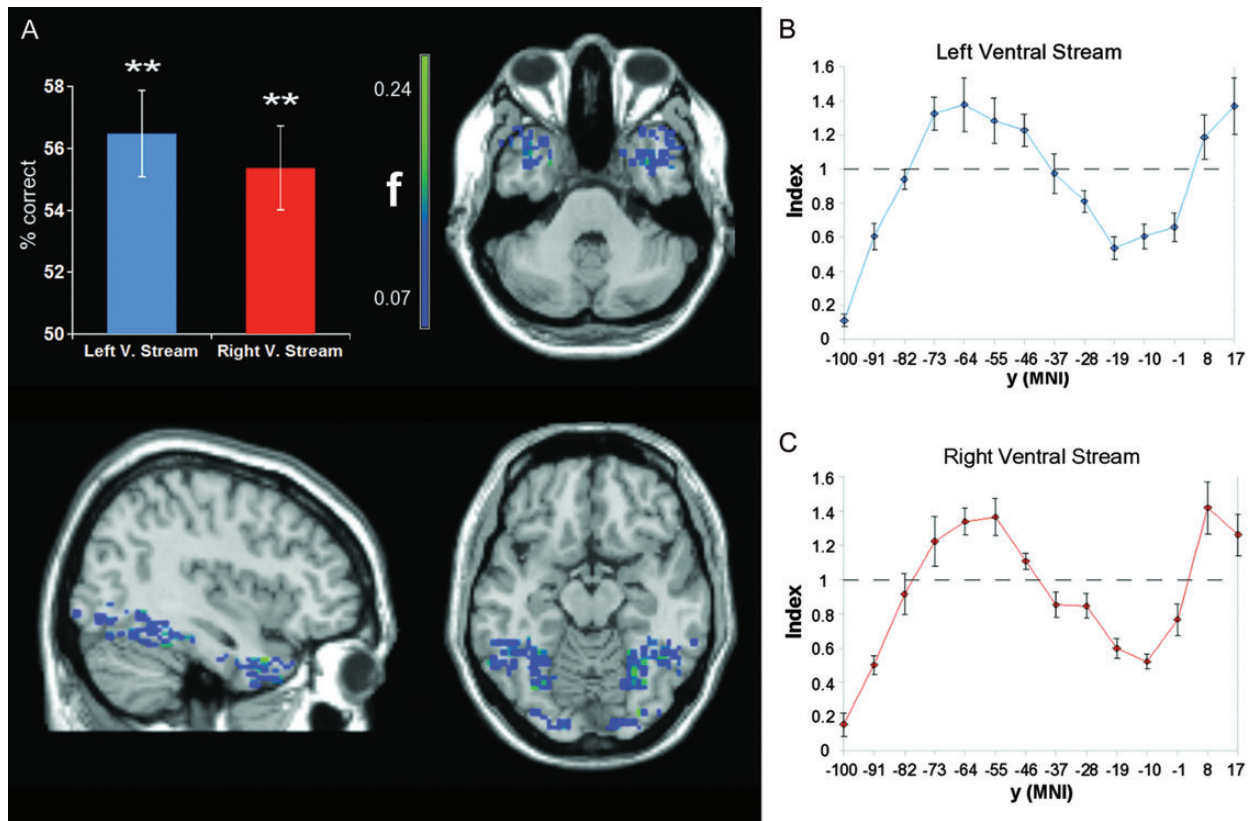


Figure 2. (A) SVM-based multivariate analysis reveals the presence of orientation-tolerant information about individual faces in the ventral stream in both hemispheres (left: accuracy = 56.48%, $t_{(8)} = 4.62$, $P < 0.005$; right: accuracy = 55.37%, $t_{(8)} = 4.56$, $P < 0.005$). The distribution of orientation-tolerant information in the ventral stream is illustrated with maps depicting the group-averaged frequency (f) with which each voxel was selected as 1 of the 200 optimal voxels for classification (separately for the 2 hemispheres). Maps are thresholded at the frequency with which a voxel would be selected if the optimal voxels were chosen randomly with a uniform distribution: (number of voxels selected)/(number of voxels in the mask). Error bars represent the standard error of the mean. $**P < 0.005$. (B) The ventral stream was divided into bins extending over a span of 3 voxels along the posterior to anterior dimension. The plots show an index given by the average probability with which voxels in a bin were among the 200 optimal voxels divided by the probability expected from a uniform distribution.

Results

ROI Analysis

Spherical ROIs for the right OFA, FFA, and ATL were generated, centered in the peaks of activity for faces in an independent functional localizer. The effect of sphere radius on classification performance was investigated testing classification with radii of 6, 9, and 12 mm. For each radius, a V1 ROI matched in number of voxels to the other ROIs was anatomically defined using MARSBAR's AAL (Tzourio-Mazoyer et al. 2002). In a first analysis, a linear SVM was trained to distinguish all views of one identity from all views of another identity on part of the data, and then tested on the discrimination of the responses to the same stimuli on the remaining data (Fig. 1B). As in the analysis performed by Nestor et al. (2011), separate data were used for training and testing, but the set of stimuli that elicited the patterns used for training and the set of stimuli that elicited the patterns used for testing are the same. Accuracy was significantly above chance for all sphere sizes (6 mm: $t_{(8)} = 2.4346$; $P < 0.05$; 9 mm: $t_{(8)} = 2.3196$; $P < 0.05$; 12 mm: $t_{(8)} = 5.5843$, $P < 0.005$) in the right ATL, only for spheres of 9 and 12 mm (6 mm: $t_{(8)} = 1.2429$; $P > 0.1$; 9 mm: $t_{(8)} = 3.5365$; $P < 0.05$; 12 mm: $t_{(8)} = 4.1$; $P < 0.005$, respectively) in FFA, and only for the 12 mm sphere ($t_{(8)} = 5.4913$; $P < 0.005$) in OFA. However, classification

performance was also significantly above chance in V1 for all sphere sizes (6 mm: $t_{(8)} = 3.1508$; $P < 0.05$; 9 mm: $t_{(8)} = 2.8642$; $P < 0.05$; 12 mm: $t_{(8)} = 2.4259$; $P < 0.05$). As it is unlikely that V1 stores orientation-tolerant representations of faces, these results suggest that this type of test is not sufficiently stringent to investigate the tolerance of face representations.

As a more stringent test, we trained a linear SVM to distinguish between 2 faces in 4 orientations, and tested whether it could classify the faces in the remaining orientation (Fig. 1C). This generalization analysis differs from the previous analysis in that the set of stimuli that elicited the patterns used for training is different from the set of stimuli that elicited the patterns used for testing. For all sphere sizes, orientation-invariant classification in the right ATL was highly significant (6 mm: $t_{(8)} = 5.94$; $P < 0.0005$; 9 mm: $t_{(8)} = 4.22$; $P < 0.005$; 12 mm: $t_{(8)} = 4.63$; $P < 0.005$). In FFA and OFA, classification accuracy was non-significant for spheres of radii 6 and 9 mm (FFA 6 mm: $t_{(8)} = 1.54$; $P > 0.05$; FFA 9 mm: $t_{(8)} = 1.96$, $P > 0.05$; OFA 6 mm: $t_{(8)} = 1.47$; $P > 0.05$; OFA 9 mm: $t_{(8)} = 2.00$, $P > 0.05$), but became significant for spheres of radius 12 mm (FFA: $t_{(8)} = 3.53$; $P < 0.01$; OFA: $t_{(8)} = 4.00$, $P < 0.005$). Importantly, orientation-invariant classification in V1 was at chance for all ROI sizes. As an additional control, orientation-invariant classification was tested in a ROI containing all visually responsive V1 voxels. The ROI was generated

selecting all voxels within the V1 ROI of MARSBAR's AAL (Tzourio-Mazoyer et al. 2002) which responded more to faces than to baseline ($P < 0.05$ uncorrected). Orientation-invariant classification in this ROI was non-significant (mean accuracy: 52.5%, $t_{(8)} = 1.49$, $P > 0.1$). To complement the SVM analysis testing for information about face identity, we studied the effects of orientation averaging the responses to different individual faces and calculating correlation matrices between the patterns of response to faces seen in different orientations. Increasing dissimilarity in the patterns of response with an increase in the rotation angle was found in V1, but not in more anterior regions (Supplementary Fig. 3).

Intracranial recordings in humans have revealed neurons in the hippocampus that respond to images of individual people showing generalization across changes in the low-level properties of the images (Quiroga et al. 2005). To test the information content about individual faces in the hippocampus, we used anatomically defined ROIs from the Wake Forest University PickAtlas IBASPM 116 library, and we performed the same stringent test applied to the other ROIs. Significant orientation-tolerant information about individual faces was detected in the hippocampus bilaterally (left hemisphere: accuracy = 54.12%, $t_{(8)} = 2.86$, $P < 0.05$; right hemisphere: accuracy = 54.86%, $t_{(8)} = 3.24$, $P < 0.05$), consistent with the reports by Quiroga et al. (2005).

Orientation-Invariant Information in the Ventral Stream

Using the more stringent generalization test, we further investigated orientation-tolerant representations of face identity in the ventral stream. The 200 most informative voxels were individuated with RFE (Guyon et al. 2002; De Martino et al. 2008), separately for the left and right hemispheres. Using activity in these voxels, linear SVMs achieved significant orientation-tolerant classification of individual faces (left ventral stream: accuracy = 56.48%, $t_{(8)} = 4.62$, $P < 0.005$; right ventral stream: accuracy = 55.37%, $t_{(8)} = 4.56$, $P < 0.005$). A map showing the probability of each voxel to be among the most informative (Fig. 2A) reveals an occipitotemporal and an anterior temporal cluster of informative voxels in each hemisphere. To provide a more quantitative evaluation of the greater concentration of informative voxels at specific locations along the posterior to anterior axis, we subdivided the ventral stream into a set of bins along the posterior to anterior axis, and we analyzed the probability of finding informative voxels in each bin. This analysis shows a greater concentration of informative voxels at the approximate y coordinates of OFA/FFA and the anterior temporal lobe (Fig. 2B). An analogous pattern is obtained when regressing out the temporal signal-to-noise ratio (TSNR) to account for differences in the quality of the BOLD signal at different levels in the ventral stream (see Supplementary Fig. 4).

To compare the classification performance obtained with each cluster to the performance obtained with the totality of the 200 voxels, we calculated an index (ratio index) given by the ratio between the above-chance accuracy obtained in a cluster (a_c) and the above-chance accuracy obtained with the total set of 200 voxels (a_{200}): $(a_c - 50\%) / (a_{200} - 50\%)$ (Fig. 3A). Significant classification was obtained in all clusters except the left anterior cluster (ratio index values: left anterior: 0.22, $t_{(8)} = 1.48$, $P = 0.09$; right anterior: 0.62, $t_{(8)} = 3.34$,

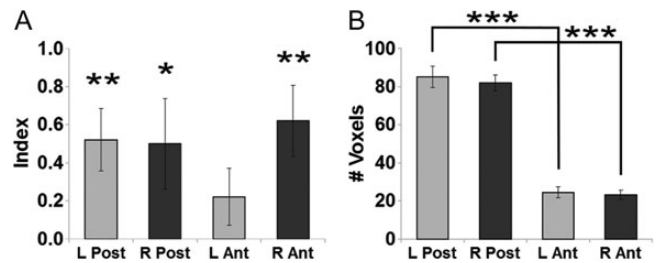


Figure 3. (A) The ratio of the above chance classification accuracy shown in Figure 2 (ratio index) is shown for the 4 clusters considered individually. As the bars denote a ratio of the above chance classification accuracy, chance level is not at 0.5 but at 0. All clusters showed above chance classification, which reached significance in the right anterior cluster ($P < 0.01$), in the left posterior cluster ($P < 0.01$), and in the right posterior cluster ($P < 0.05$), and trended to significance in the left anterior cluster ($P = 0.09$). $**P < 0.01$, $*P < 0.05$. (B) Number of voxels in the individual clusters. Anterior clusters contain a smaller number of voxels than posterior clusters. $***P < 0.0001$. Error bars represent the standard error of the mean.

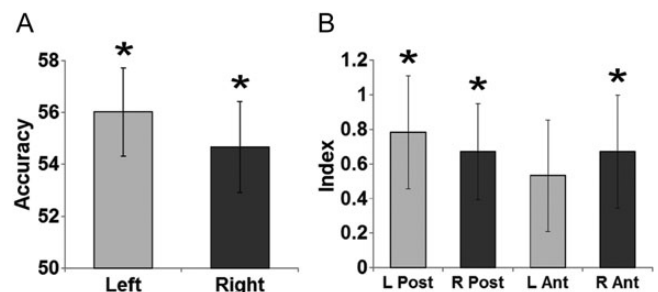


Figure 4. (A) Tolerant classification in the ventral stream of the 2 hemispheres when excluding from the training set the views that are mirror symmetrical to the testing views; (B) ratio index for the 4 clusters of informative voxels when excluding the mirror symmetrical views from the training set. Error bars represent the standard error of the mean.

$P < 0.01$; left posterior: 0.52, $t_{(8)} = 3.20$, $P < 0.01$; right posterior: 0.5, $t_{(8)} = 2.10$, $P < 0.05$, 1-tailed tests). Classification accuracy in these clusters was slightly lower than in the spherical ROIs (Fig. 1C). This is probably due to the smaller number of voxels used for classification in this analysis, about 20 in anterior regions and 80 in posterior regions, when compared with the 51 for 6-mm radius spheres and 381 for 12-mm radius spheres.

For each participant, we calculated the average number of informative voxels in the posterior and anterior clusters. The average number of voxels in anterior clusters is significantly lower than the average number of voxels in posterior clusters (left: $t_{(8)} = 7.45$, $P < 0.0001$; right: $t_{(8)} = 9.69$, $P < 0.0001$, Fig. 3B). Given the comparable accuracy obtained with the posterior clusters and the right anterior cluster, this indicates that there is greater information per mm^3 of cortex in the right ATL than in the 2 occipitotemporal clusters.

Effects of Mirror Symmetry

In the previous analyses, 4 orientations of the faces were used for training the classifier, and the remaining orientation was used for testing. Whenever the orientation used for testing was not the frontal orientation, the data used for training included images that are the mirror reflection of the images used for testing (e.g., if the testing orientation was -35° , the training orientations included the $+35^\circ$ orientation). If

representations in a brain region were tolerant for mirror reflections, the tolerant classification observed might entirely depend on the classifier exploiting the similarity between the response to the testing image and the response to its mirror reflection present in the training set. To test this possibility, we repeated the analysis excluding from the training set the orientations that are mirror reflections of the testing orientations. Classification remained significantly above chance in both hemispheres (left: 56.02%, $t_{(8)} = 3.5423$, $P < 0.05$; right: 54.67%, $t_{(8)} = 2.6681$, $P < 0.05$, see Fig. 4A). When considering the individual clusters separately, classification accuracy is significantly above chance in all clusters except the left anterior cluster, as in the case of the analysis that included the mirror symmetrical orientations. The ratio index ranges from 0.53 to 0.78 (left posterior: 0.78, $t_{(8)} = 2.41$, $P < 0.05$; right posterior: 0.67, $t_{(8)} = 2.41$, $P < 0.05$; left anterior: 0.53, $t_{(8)} = 1.65$, n.s.; right anterior: 0.67, $t_{(8)} = 2.05$; $P < 0.05$; 1-tailed t -tests, see Fig. 4B).

Discussion

Several fMRI studies investigated specificity and tolerance in the representation of individual faces and other objects using fMRI adaptation (fMR-A), with mixed results: some studies found evidence for adaptation (Ewbank and Andrews 2008; Mur et al. 2010) while others did not (Pourtois et al. 2005). More importantly, a recent study (Mur et al. 2010) found lower BOLD signal in early visual cortex in response to faces for which a different view had been previously presented than for novel faces. Given the current understanding of representations in early visual cortex, these results suggest that finding adaptation effects for different orientations of an object in a brain region does not necessarily imply that that region encodes orientation-tolerant representations of that object.

In this study, we showed that using the patterns of BOLD signal in the ventral stream it is possible to classify individual faces, generalizing across rotations in depth. Highly informative voxels for this classification cluster in the anterior temporal lobes and the ventral occipitotemporal cortex. Generalization accuracy obtained with each region in isolation is comparable (with the exception of the left ATL). However, fewer informative voxels were individuated in anterior regions. Within standard face-responsive ROIs, significant orientation-tolerant classification was found in the right ATL, FFA, and OFA, while orientation-tolerant classification in V1 was at chance. These results show that occipitotemporal cortex and the ATL do not just represent specific images of faces but the identity of a face with tolerance for changes in orientation. This is consistent with reports of deficits for face recognition following damage to the occipitotemporal cortex (Meadows 1974; De Renzi et al. 1994) and to the ATL (Evans et al. 1995; Tranel et al. 1997).

An interesting aspect of the present results is that we found orientation-tolerant representations of faces in both occipitotemporal cortex and ATL. This finding raises the question of what are the respective roles of these regions for invariant recognition of individual faces. Comparable classification accuracy does not imply that the right anterior cluster and the 2 posterior clusters store representations of the same type: posterior regions might store representations that carry a greater amount of information about perceptual details of the faces, while the right ATL might represent the identity of the faces

abstracting away from perceptual details. The difference in the number of informative voxels between anterior and posterior regions suggests some tentative conclusions about their respective roles in face recognition. If representations in the human ATL abstract away from the perceptual details of faces (in line with single-cell recording studies in monkeys that show greater tolerance for rotations in anterior regions of the temporal lobe, [Freiwald and Tsao 2010](#)) less information would have to be represented, and it could be represented over a smaller extent of cortex. Therefore, the present results are consistent with the possibility that representations in anterior regions abstract away from the perceptual details of the images to a greater extent than representations in posterior regions. In this respect, the finding of significant orientation-tolerant classification as early as in the OFA is of particular interest, especially in the context of recent results suggesting that activity in the OFA represents face parts without being modulated by their configuration ([Liu et al. 2010](#)). Taken together, these results suggest that some degree of invariance across different orientations could be achieved at the level of representations of face parts, without necessarily implying face/identity recognition.

Our findings are consistent with models of face recognition proposing that faces are processed by a cortical network with some regions encoding static aspects of faces such as identity and other regions encoding changeable aspects of faces such as viewpoint and expression ([Haxby et al. 2000](#); [Ishai 2008](#)). In particular, our results are in line with the view that static aspects of faces are processed in ventral temporal cortex. Furthermore, the processing of static aspects of faces itself seems to be subserved by a network comprising cortical regions in posterior and anterior portions of the ventral stream. In addition to these neocortical areas, tolerant representations of face identity were also found in the hippocampus, consistent with electrophysiological studies in humans ([Quiroga et al. 2005](#)). Given the involvement of the hippocampus in episodic memory ([Vargha-Khadem et al. 1997](#)), these representations might play an important role in the association between a person's appearance and what we remember about previous interactions with that person. This kind of associations could enrich our knowledge of a person's identity beyond its physical appearance.

It is important to note that despite showing some degree of orientation tolerance, representations of faces in the occipitotemporal cortex are not sufficient for normal face recognition. Patients with ATL damage and intact occipitotemporal cortices can show marked face recognition impairments ([Warrington and Shallice 1984](#); [Tyrrell et al. 1990](#); [Evans et al. 1995](#)). Furthermore, [Avidan and Behrmann \(2009\)](#) report normal repetition suppression effects for faces in the FFA of a group of subjects affected by congenital prosopagnosia, and [Thomas et al. \(2009\)](#) found that the anatomical connectivity between the FFA and the ATL (as measured with diffusion tensor imaging) is reduced in congenital prosopagnosics, providing additional evidence for the importance of the ATL for face recognition.

Informative voxels were found with RFE also outside the functionally defined ROIs. Orientation-tolerant information about individual faces might be present in these voxels because they contain neurons involved in the recognition of nonface objects that share some similarities in shape with faces or face parts. Alternatively, these voxels might contain

small populations of neurons involved in face recognition that are located outside the OFA, FFA, and ATL face-responsive ROIs.

Single-cell physiology studies have reported neurons responding to mirror symmetrical views of objects (Logothetis et al. 1995; Freiwald and Tsao 2010), and a recent study (Dilks et al. 2011) reported adaptation effects for mirror symmetrical views of faces. Therefore, we tested whether the observed tolerance in our study depends entirely on the presence in the training set of responses to images that are mirror reflections of the images in the testing set. We found significantly above chance orientation-tolerant classification even when the responses to mirror reflections of images in the testing set were excluded from the training set. Therefore, the observed tolerance cannot be explained by the similarity between the responses to mirror symmetrical views. This finding does not imply that there is no effect of mirror symmetry on the similarity between the patterns of responses in the regions investigated, it only implies that the representations in these regions are also tolerant for image changes other than mirror reflections.

There is a substantial body of evidence suggesting that the ATL is involved in the representation of social knowledge (Simmons et al. 2010; Zahn et al. 2007, see Olson et al. 2007 for a review), suggesting to some that this region is specialized for representing this type of knowledge (e.g., Simmons et al. 2010). However, in this study, we found that the ATL contains information that is sufficient for the classification of individual faces. This despite the fact that faces were of the same gender and ethnicity, and no names or biographical facts were associated with the faces. This finding suggests that the ATL is not exclusively involved in representing semantic facts about people and groups. Representations of individual faces near or spatially overlapping with representations of social knowledge might provide the neural basis of our ability to associate social knowledge with perceptual experience.

In a previous study (Anzellotti et al. 2011), we found that distinct subregions within the ATL responded to animals and tools, respectively. Taken together, these results show that despite the fact that parts of the ATL are activated during the processing of social knowledge, ATL as a whole is not exclusively involved in the processing of social knowledge. This does not exclude the possibility that some subregion of the ATL might be specialized for processing social knowledge. To address this question, it will be important to test multiple contrasts within the same participants, investigating whether and to what extent the portions of the right ATL that are activated by social knowledge also respond to other kinds of objects (e.g., animals) or contain information that allows classification of stimuli that do not have any associated social knowledge.

In future studies, it will be interesting to investigate the correlation between behavioral performance of individual participants and classification performance in different brain regions, and to study whether individual differences in the patterns of brain activity in the ventral stream predict behavioral differences. Furthermore, it would be possible to test whether differences in the patterns of brain activity allow to predict within-participant differences in the discrimination performance between pairs of faces. Some participants might find it easier to discriminate a face A from a face B than A from C, and others might show the opposite behavioral pattern. Investigating where in the brain we can find

differences in the classification between individual faces that correlate with the individuals' face recognition performance would help to clarify the role played by different brain regions for face recognition.

Another line of research that would be worth pursuing consists in investigating how other kinds of image transformations affect the neural representations of faces. For example, it would be possible to study where in the brain there are representations of face identity that generalize across changes in illumination, and whether the same voxels that contribute to illumination-tolerant classification also allow orientation-tolerant classification.

In conclusion, the present study shows that occipitotemporal cortex and ATL do not just represent specific images of faces but represent also identity information with tolerance for image transformations. Furthermore, the tolerance observed in these regions cannot be explained by the similarity of the neural responses to mirror symmetrical views. Representations in the right anterior temporal lobe are more "compact" than representations in posterior regions, and might abstract away from the perceptual details of the images. Interestingly, orientation-invariant classification is obtained as early as in the OFA, suggesting the presence of part-based mechanisms for orientation-invariant recognition.

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Funding

This work was supported by the Provincia Autonoma di Trento and by the Fondazione Cassa di Risparmio di Trento e Rovereto.

Notes

We thank Yaoda Xu and Ken Nakayama for advice on the analyses, Jorge Jovicich and Simon Robinson for the development of the EPI sequence, Claudio Boninsegna, Fabrizio Pallaver, Hanna Liisa Inkala, and Manuela Orsini for technical assistance with the MR. Stefano Anzellotti was supported by a grant from the Fondazione Cassa Rurale di Trento. *Conflict of Interest:* None declared.

References

- Anzellotti S, Mahon BZ, Schwarzbach J, Caramazza A. 2011. Differential activity for animals and manipulable objects in the anterior temporal lobes. *J Cog Neurosci*. 23(8):2059–2067.
- Avidan G, Behrmann M. 2009. Functional MRI reveals compromised neural integrity of the face processing network in congenital prosopagnosia. *Curr Biol*. 19:1146–1150.
- Boser BE, Guyon IM, Vapnik V. 1992. A training algorithm for optimal margin classifiers. Fifth Annual Workshop on Computational Learning Theory, Pittsburgh. ACM.
- Brainard DH. 1997. The psychophysics toolbox. *Spat Vis*. 10:443–446.
- Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of interest analysis using an SPMtoolbox [abstract]. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. *Neuroimage*. 16(2): Elsevier.
- Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR. 1996. A neural basis for lexical retrieval. *Nature*. 380:499–505.
- De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E. 2008. Combining multivariate voxel selection and support

- vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*. 43:44–58.
- De Renzi E, Perani D, Carlesimo GA, Silveri MC, Fazio F. 1994. Prosopagnosia can be associated with damage confined to the right hemisphere—an MRI and PET study and a review of the literature. *Neuropsychologia*. 32:893–902.
- Dilks DD, Julian JB, Kubilius J, Spelke ES, Kanwisher N. 2011. Mirror-image sensitivity and invariance in object and scene processing pathways. *J Neurosci*. 31(31):11305–11312.
- Evans JJ, Heggs AJ, Antoun N, Hodges JR. 1995. Progressive prosopagnosia associated with selective right temporal lobe atrophy: a new syndrome? *Brain*. 118:1–13.
- Ewbank MP, Andrews TJ. 2008. Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. *Neuroimage*. 40:1857–1870.
- Freiwald WA, Tsao DY. 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*. 330:845–851.
- Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW. 2000. The fusiform “face area” is part of a network that processes faces at the individual level. *J Cogn Neurosci*. 12:495–504.
- Guyon I, Weston J, Barnhill S, Vapnik V. 2002. Gene selection for cancer classification using support vector machines. *Machine Learn*. 46:389–422.
- Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn Sci*. 4:223–233.
- Ishai A. 2008. Let's face it: it's a cortical network. *Neuroimage*. 40:415–419.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 17:4302–4311.
- Kriegeskorte N, Formisano E, Sorger B, Goebel R. 2007. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc Natl Acad Sci USA*. 104:20600–20605.
- 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, Poggio T. 1995. “Shape representation in the inferior temporal cortex of monkeys”. *Curr Biol*. 5(5): 552–563.
- Lopes da Silva FH, Arnolds DEAT. 1978. Physiology of the hippocampus and related structures. *Annu Rev Physiol*. 40(1): 185–216.
- Meadows JC. 1974. The anatomical basis of prosopagnosia. *J Neurol Neurosurg Psychiatry*. 37:489–501.
- Mur M, Ruff DA, Bodurka J, Bandettini PA, Kriegeskorte N. 2010. Face-identity change activation outside the face system: “Release from adaptation” may not always indicate neuronal selectivity. *Cereb Cortex*. 20(9):2027–2042.
- 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.
- Olson IR, Plotzker A, Ezzyat Y. 2007. The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*. 130:1718–1731.
- Pelli DG. 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*. 10:437–442.
- Pourtois G, Schwartz S, Seghier ML, Lazeyras F, Vuilleumier P. 2005. View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: an event-related fMRI study. *Neuroimage*. 24:1214–1224.
- Puce A, Allison T, Asgari M, Gore JC, McCarthy G. 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci*. 16:5205–5215.
- Puce A, Allison T, Gore JC, McCarthy G. 1995. Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J Neurophysiol*. 74:1192–1199.
- Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. 2005. Invariant visual representation by single neurons in the human brain. *Nature*. 435:1102–1107.
- Rajimehr R, Young JC, Tootell RBH. 2009. An anterior temporal face patch in human cortex, predicted by macaque maps. *Proc Natl Acad Sci USA*. 106(6):1995–2000.
- Schwarzbach J. 2011. A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behav Res Methods*.
- Sergent J, Ohta S, MacDonald B. 1992. Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain*. 115:15–36.
- Simmons WK, Reddish M, Bellgowan PSF, Martin A. 2010. The selectivity and functional connectivity of the anterior temporal lobes. *Cereb Cortex*. 20:813–825.
- Thomas C, Avidan G, Humphreys K, Jung KJ, Gao F, Behrmann M. 2009. Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia. *Nat Neurosci*. 12:29–31.
- Tranel D, Damasio H, Damasio AP. 1997. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*. 35:1319–1327.
- Tyrrell PJ, Warrington EK, Frackowiak RSJ, Rossor MN. 1990. Progressive degeneration of the right temporal lobe studied with positron emission tomography. *J Neurol Neurosurg Psychiatry*. 53:1046–1050.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labelling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single subject brain. *Neuroimage*. 15:273–289.
- Vargha-Khadem F, Gadian DG, Watkins KE, Connelly A, Van Paesschen W, Mishkin M. 1997. Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*. 277 (5324):376–380.
- Vetter T, Poggio T, Bulthoff HH. 1994. The importance of symmetry and virtual views in three-dimensional object recognition. *Curr Biol*. 4:18–23.
- Warrington EK, Shallice T. 1984. Category-specific semantic impairments. *Brain*. 107:829–854.
- Zahn R, Moll J, Krueger F, Huey ED, Garrido G, Grafman J. 2007. Social concepts are represented in the superior anterior temporal cortex. *Proc Natl Acad Sci USA*. 104:6430–6435.