

Perception and Processing of Faces in the Human Brain Is Tuned to Typical Feature Locations

 Benjamin de Haas,^{1,2,3} D. Samuel Schwarzkopf,^{1,3}  Ivan Alvarez,^{4,5}  Rebecca P. Lawson,^{1,2}  Linda Henriksson,^{6,7} Nikolaus Kriegeskorte,⁶ and  Geraint Rees^{1,2}

¹Institute of Cognitive Neuroscience, ²Wellcome Trust Centre for Neuroimaging, ³Experimental Psychology, and ⁴Institute of Child Health, University College London, London WC1H 0AP, United Kingdom, ⁵Oxford University Centre for Functional MRI of the Brain, Oxford OX3 9DU, United Kingdom, ⁶MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, United Kingdom, and ⁷Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo FI-00076, Finland

Faces are salient social stimuli whose features attract a stereotypical pattern of fixations. The implications of this gaze behavior for perception and brain activity are largely unknown. Here, we characterize and quantify a retinotopic bias implied by typical gaze behavior toward faces, which leads to eyes and mouth appearing most often in the upper and lower visual field, respectively. We found that the adult human visual system is tuned to these contingencies. In two recognition experiments, recognition performance for isolated face parts was better when they were presented at typical, rather than reversed, visual field locations. The recognition cost of reversed locations was equal to ~60% of that for whole face inversion in the same sample. Similarly, an fMRI experiment showed that patterns of activity evoked by eye and mouth stimuli in the right inferior occipital gyrus could be separated with significantly higher accuracy when these features were presented at typical, rather than reversed, visual field locations. Our findings demonstrate that human face perception is determined not only by the local position of features within a face context, but by whether features appear at the typical retinotopic location given normal gaze behavior. Such location sensitivity may reflect fine-tuning of category-specific visual processing to retinal input statistics. Our findings further suggest that retinotopic heterogeneity might play a role for face inversion effects and for the understanding of conditions affecting gaze behavior toward faces, such as autism spectrum disorders and congenital prosopagnosia.

Key words: decoding; face perception; fMRI; gaze behavior; occipital face area; retinotopy

Significance Statement

Faces attract our attention and trigger stereotypical patterns of visual fixations, concentrating on inner features, like eyes and mouth. Here we show that the visual system represents face features better when they are shown at retinal positions where they typically fall during natural vision. When facial features were shown at typical (rather than reversed) visual field locations, they were discriminated better by humans and could be decoded with higher accuracy from brain activity patterns in the right occipital face area. This suggests that brain representations of face features do not cover the visual field uniformly. It may help us understand the well-known face-inversion effect and conditions affecting gaze behavior toward faces, such as prosopagnosia and autism spectrum disorders.

Introduction

Traditionally, visual object recognition is described as a location-invariant process, relying on neurons of the inferior temporal

cortex that prefer objects of a particular category while being largely insensitive to stimulus position (Riesenhuber and Poggio, 1999). However, accumulating evidence points to a role for stimulus location in object perception (Afraz et al., 2010) and object selective areas of visual cortex (Kravitz et al., 2008, 2010; Schwarzlose et al., 2008; Chan et al., 2010; Kay et al., 2015; Silson et al., 2015). Activity in human object-selective cortex exhibits a link

Received Oct. 6, 2014; revised May 6, 2016; accepted May 15, 2016.

Author contributions: B.d.H., D.S.S., L.H., N.K., and G.R. designed research; B.d.H., I.A., and R.P.L. performed research; L.H. and N.K. contributed unpublished reagents/analytic tools; B.d.H. analyzed data; B.d.H., D.S.S., I.A., R.P.L., L.H., N.K., and G.R. wrote the paper.

This work was supported by the Wellcome Trust to B.d.H., D.S.S., R.P.L., and G.R., Deutsche Forschungsgemeinschaft research fellowship HA 7574/1-1 to B.d.H., European Research Council Starting Grant 310829 to D.S.S. and B.d.H., European Research Council Starting Grant 261352 to N.K., Wellcome Trust Project Grant WT091540MA to N.K., Medical Research Council to L.H. and N.K., and Academy of Finland Postdoctoral Grant 278957 to L.H. The Wellcome Trust Centre for Neuroimaging is supported by core funding from the Wellcome Trust 091593/Z/10/Z/.

The authors declare no competing financial interests.

This article is freely available online through the *JNeurosci* Author Open Choice option.

Correspondence should be addressed to Dr. Benjamin de Haas, Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom. E-mail: benjamin.haas.09@ucl.ac.uk.

DOI:10.1523/JNEUROSCI.4131-14.2016

Copyright © 2016 de Haas et al.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

between eccentricity and category preferences (Hasson et al., 2002; Konkle and Oliva, 2012) as well as a spatial preference for face and body halves appearing at their typical, contralateral side of the visual field (Chan et al., 2010). In macaques, cells of the posterior lateral face patch show a preference for stimuli depicting the contralateral eye as well as for the contralateral upper visual field quadrant (Issa and DiCarlo, 2012). This raises the question whether tuning for facial features and for location are more generally matched to each other. Specifically, does such a match extend to facial features other than the eye, is it found in humans and does it have perceptual relevance?

Face perception is an important example of category-specific visual processing in the human visual system. Especially the eye and mouth regions are important for face recognition and categorization, as well as for neural responses to faces (e.g., Smith et al., 2004, 2009). Moreover, recognition of faces and face features is determined by the relative arrangement of features within a face context (e.g., Tanaka and Farah, 1993; Tanaka and Sengco, 1997; Maurer et al., 2002). However, it is an open question whether face perception is tuned to absolute, retinotopic feature locations. Importantly, human observers show a stereotypical pattern of gaze behavior toward faces. Typically, first fixations land on the central upper nose region, just below the eyes (Hsiao and Cottrell, 2008; Peterson and Eckstein, 2012), and subsequent fixations remain restricted to inner features (e.g., van Belle et al., 2010; Vö et al., 2012). This pattern of gaze behavior implies a retinotopic bias: eyes will appear more often in the upper than lower visual field and vice versa for mouths. Recent evidence suggests a map-like organization of face feature preferences in the occipital face area (Henriksson et al., 2015). This result has been interpreted to stem from typical gaze behavior: A retinotopic cortical proto-map might acquire tuning to eyes, mouths, and other face features in the positions most frequently exposed to these features, forming a “faciotopic” map (Henriksson et al., 2015; van den Hurk et al., 2015).

Here, we explicitly quantified the retinotopic bias and found that recognition and neural processing of facial features is tuned to canonical visual field locations. Tracking the gaze of observers freely viewing faces we found that, for most fixations, eyes and mouths appeared in the upper and lower visual field, respectively. Crucially, recognition performance for individuating isolated eye and mouth images showed a strong feature \times location interaction. Observers performed significantly better when features were presented at typical visual field locations. An independent recognition experiment replicated this effect in a larger sample and showed that the recognition cost of reversing feature locations, summed across eyes and mouth, amounted to 60%–80% of the effect size for whole face inversion. Finally, in an fMRI experiment, we probed the ability to decode eye- versus mouth-evoked patterns of activity from face-sensitive areas of the ventral stream. Decoding yielded significantly higher accuracies for typical rather than reversed visual field locations in right inferior occipital gyrus (IOG). This suggests that feature and location tuning of the underlying neural populations may be matched.

Materials and Methods

Gaze experiment

Participants. Fourteen healthy participants from the University College London (UCL) participant pool took part in the gaze experiment (age 21–60 years, mean 39 years, SD 14.01 years; 9 females). All participants had normal or corrected-to-normal vision and were right-handed. Each participant provided written informed consent, and the experiments were approved by the UCL Research Ethics Committee.

Stimuli. A set of 24 grayscale frontal photographs of faces with neutral expression was taken from an online face database (Minear and Park, 2004) (<http://agingmind.utdallas.edu/facedb>). Images were rectangular with an onscreen width and height of 29.51×21.92 cm at a viewing distance of 80 cm (20.90×15.60 degrees visual angle). The horizontal (ear to ear) and vertical (hairline to chin) onscreen size of faces contained in the images ranged from 8.5 to 11.0 cm (6.1–7.9 degrees visual angle) and from 12.5 to 15.5 cm (8.9–11.1 degrees visual angle), respectively. Images were presented on a gray background on a liquid crystal display monitor (LCD; Samsung SyncMaster 2233RZ) with a refresh rate of 120 Hz and a spatial resolution of 1680×1050 pixels.

Procedure. Participants sat on a chair with their head in a chin rest at a viewing distance of 80 cm. They were asked to “look at the faces in any manner [they] wish” and started the consecutive presentation of 12 faces with a button press on a keyboard. The order of faces was randomized for each participant, and each face was shown for 2000 ms followed by a 2000 ms gray screen. Each participant saw two blocks of 12 faces. Gaze direction was monitored with an infrared eyetracker (Cambridge Research Systems) tracking the left eye at 200 Hz.

Analysis. Gaze direction could successfully be tracked for an average of 80.04% of samples ($\pm 5.93\%$ SEM). From these, we excluded any samples falling outside the images ($0.24 \pm 0.15\%$ SEM) or classified as saccades (based on a velocity criterion of >30 deg/s or 0.15 deg between successive samples; $34.22 \pm 6.51\%$ SEM). Exclusion of these samples did not change the pattern of results qualitatively. To plot visual field positions of facial features for the remaining fixations, we calculated the position of mouth and eye features relative to respective fixation samples. For this, we manually determined the center position of the mouth and either eye for each face image.

Recognition experiments

Participants. Separate samples of healthy participants from the UCL participant pool took part in two recognition experiments. All participants in Experiments 1 ($n = 18$; age 19–52 years, mean 27 years, SD 9 years; 11 females; 3 left-handed) and 2 ($n = 36$; age 19–37 years, mean 25 years, SD 5 years; 26 females; 2 left-handed) had normal or corrected-to-normal vision. Written informed consent was obtained from each participant, and the experiments were approved by the UCL Research Ethics Committee.

Stimuli. Face feature stimuli stem from a set of 54 frontal photographs of faces with neutral expression, none of which contained easily identifiable eye or mouth features, such as facial hair (Henriksson et al., 2015). These images were used to form 27 face candidate pairs matched for gender, skin color, and eye color (dark/light).

Each of the 27 candidate faces yielded a set of three candidate face features (left eye, right eye and mouth images, respectively). These face feature images were sampled according to a symmetric grid of squares that was overlaid on face images, such that two squares were centered on the left and right eye region and one on the mouth region. The respective tile regions were cut out and served as feature images in the experiment. Feature images were square and presented at a visual angle of 3.5 and 4 degrees in Experiments 1 and 2, respectively.

The stimulus set of Experiment 2 additionally contained 54 whole face images, which were selected from the SiblingsDB set (Vieira et al., 2013) (model age restricted to 14–28 years; database available upon request to andrea.bottino@polito.it). These images were realigned, cropped to a square with edges just below the lower lip and just above the eyebrows, and scaled to 11 degrees visual angle. Whole face images were matched as nonsibling pairs based on gender, skin, and eye color and had a maximum age difference of 7 years.

The outer edge of each image was overlaid with a gray fringe that softened the edge between image and background and was ~ 0.5 and ~ 1 degrees wide for feature and whole face images, respectively. Participants saw 8-bit grayscale versions of the images that were displayed on a gray background and with a dynamic noise mask overlay (see below). Stimuli were shown on a liquid crystal display monitor (LCD; Samsung SyncMaster 2233RZ) with a refresh rate of 120 Hz and a spatial resolution of 1680×1050 pixels.

Procedure. Participants sat on a chair with their head in a chin rest at a viewing distance of 80 cm (Experiment 1) or 44 cm (Experiment 2). Each

trial began with the presentation of a blue fixation dot at the middle of the screen and on gray background (~ 0.1 degrees visual angle in diameter). After 500 ms, the image of a face feature or whole face flashed up for 200 ms, overlaid by a dynamic noise mask that lasted until 450 ms after image onset. Each frame of the noise mask consisted of a random intensity image of the same size as the face feature images and with pixel intensities drawn from a uniform distribution ranging from 0 to 128 (corresponding to the background gray level; the luminance range of the display was not linearized). Pixel intensities of overlay frames corresponded to the numerical average of the noise mask and face feature images.

Immediately after the offset of the noise mask, the fixation dot turned green and two candidate images appeared on the screen, prompting the participant to indicate which of them flashed up earlier. Candidate images were scaled versions (75% and 100% original size in Experiments 1 and 2, respectively) of the target image and the image of the corresponding face feature in the matched candidate (if the target was a whole face, the matched candidate also was a whole face; see above). Participants used the arrow keys of a standard keyboard to shift a blue selection rectangle surrounding either candidate image and confirmed their choice with the space bar. The selection rectangle then turned either green or red for 300 ms, indicating that the answer was correct or incorrect, respectively. In Experiment 2, face feature images and whole faces were either flashed upright or inverted and the orientation of candidate images always corresponded to the preceding target.

In Experiment 1, each participant completed nine blocks of 48 trials each. A block contained trials corresponding to three sets of candidate images or six faces. Feature images of each face served as stimuli in a total of eight trials, corresponding to eight trial types. Trial types were defined by face feature (left eye, right eye, mouth) and stimulus position. There were three main stimulus positions that corresponded to the locations of the left eye, right eye, and mouth segments in the original image, assuming fixation slightly above the nose. Specifically, the mouth position was centered on the vertical meridian at 4.4 degrees visual angle below fixation, and the left and right eye positions were centered 2.6 degrees above fixation and shifted 3.5 degrees to the left or right, respectively. All center positions had an equal eccentricity of 4.4 degrees (for a schematic illustration, compare Fig. 1C).

The eight trial types corresponded to either eye image appearing at its “typical” location (two trial types corresponding to an upper visual field eye condition) or the mouth position (two trial types corresponding to a lower visual field eye condition) and the mouth image appearing at either eye position (two trial types corresponding to an upper visual field mouth condition) or the mouth image appearing at the mouth position. To balance the design, the trials of the latter trial type were repeated in each block, yielding two (dummy coded) trial types corresponding to a lower visual field mouth condition. The order of trials was randomized within each block; and in each trial, the exact stimulus position was determined as the main location (corresponding to trial type) plus a random offset to avoid adaptation or fatigue. Spatial scatter on the horizontal and vertical plane were drawn independently from a Gaussian distribution centered on zero with an SD of 0.35 degrees visual angle and clipped at 2 SDs.

The design of Experiment 2 was very similar to Experiment 1. Feature positions were scaled to reflect the slightly larger size of feature images. The mouth position was centered 5 degrees of visual angle below fixation on the vertical meridian, and the eye positions were 3 degrees above and 4 degrees to the left or right of fixation. Trials with reversed stimulus positions did not swap eye and mouth positions, as in Experiment 1, but emulated those in an inverted face; that is, upper visual field mouth trials presented the mouth at 5 degrees visual angle above fixation on the vertical meridian and lower visual field eye trials presented an eye at 3 degrees below and 4 degrees to the left or right of fixation (for a schematic illustration, see Fig. 1C).

Experiment 2 had a total of 18 different trial types. In addition to the eight trial types of the first experiment (upright features presented at typical or reversed positions), the second experiment had eight corresponding inverted trial types (inverted features presented at typical or reversed positions), as well as two trial types showing upright and inverted whole faces, respectively. The experiment was divided in four

blocks of 120 trials each, comprising of six trials for each of the 16 trial types involving feature stimuli and 12 trials each for the upright and inverted whole faces. This way, the overall number of trials was similar between Experiments 1 and 2 (432 and 480, respectively), but Experiment 2 had fewer trials per trial type (24 and 48 for feature and whole face images, respectively, versus 54 in Experiment 1). We decided to keep the overall duration of the experiment short to avoid participant fatigue (as indicated by pilot data). To compensate for potential reductions in accuracy on the single-participant level, we doubled the number of participants in Experiment 2 relative to Experiment 1.

To ensure fixation, gaze direction was monitored with an infrared eyetracker (Cambridge Research Systems) tracking the left eye at 200 Hz in both recognition experiments. Gaze data were collected for 13 participants in Experiment 1 and 33 in Experiment 2 (calibration of the eyetracker or recording failed for the remaining participants). For these participants, gaze direction could successfully be tracked for an average of 87.11% of trials (SEM = 5.29%) and 80.44% of trials (SEM = 4.83%) during Experiments 1 and 2, respectively.

Analysis. All statistical analyses of recognition data were performed in MATLAB (The MathWorks) and PASW 20 (SPSS/IBM). To test for an interaction between face feature and visual field position, the proportion of correct answers was averaged for each participant and condition: location (upper/lower visual field) by stimulus (eye/mouth). The resulting values were compared across conditions using a repeated-measures GLM and *post hoc t* tests. Additionally, we calculated the reduction in correct answers for each condition and feature, relative to its upright version shown at typical locations. This way, we quantified the recognition cost of showing a feature at its atypical location (Experiments 1 and 2), or its inverted version (Experiment 2), or the combination of inversion and atypical location (Experiment 2). For Experiment 2 we used a repeated-measures GLM with within-subject factors stimulus (eye/mouth) and manipulation (location/inversion/combination) to compare the recognition cost of these manipulations. Finally, we contrasted recognition performance for upright and inverted whole faces to quantify the recognition cost of whole face inversion and compare it with feature manipulation effects.

An indicator of general fixation compliance was computed as the median absolute deviation of gaze direction during stimulus presentation (excluding the poststimulus noise mask). This was done separately for the vertical and horizontal axes. Additionally, an indicator of gaze bias toward the stimulus was computed. For this, gaze direction on the vertical axis was compared between trials in which stimuli were presented in the upper versus lower visual field. A bias index was defined as the median difference in vertical eye position between these trial types. This bias index was also calculated separately for and compared between trials with typical versus reversed stimulus locations. Finally, to test whether lack of fixation compliance predicted the hypothesized effect, a correlation between individual effect size and propensity for eye movements was computed across participants. For this, effect size was defined as the individual recognition advantage in typical versus reversed trials and the propensity for eye movements was defined as the median absolute deviation of gaze direction during stimulus presentation (averaged across the horizontal and vertical axes).

Neuroimaging experiment

Participants. A separate sample of 21 healthy participants from the UCL participant pool took part in the fMRI study (age 20–32 years, mean 25 years, SD 4 years; 14 females; 1 left-handed). All participants had normal or corrected-to-normal vision. Written informed consent was obtained from each participant and the study was approved by the UCL Research Ethics Committee.

Stimuli. Stimuli were identical to the first recognition experiment (see above). Participants viewed stimuli via a mirror mounted on the head coil at a viewing distance of ~ 73 cm, resulting in an effective stimulus size of 3.2×3.2 degrees of visual angle.

Each participant completed an additional localizer run consisting of alternating blocks of images depicting faces and everyday objects. Face images were frontal color photographs of 10 male and 10 female faces with neutral expression taken from the PUT face database (Kasiński et al.,

2008) (<https://biometrics.cie.put.poznan.pl/>). Object images were color photographs depicting 20 everyday objects (e.g., a boot, a wall clock, or a flower). Jonas Kubilius and Dr. Lee de-Witt (KU Leuven) kindly provided these images. All localizer images were quadratic, shown on a gray background at the center of the screen and at a size of 4×4 degrees visual angle. The center of the images contained a circular cut out (width: 0.4 degrees visual angle) with a black fixation dot (width: 0.1 degrees visual angle).

All stimuli were presented with a projector at a resolution of 1280×1024 pixels and with a refresh rate of 60 Hz. Stimulus presentation was controlled using MATLAB (The MathWorks) and the Psychophysics Toolbox 3 extension (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) (<http://psychtoolbox.org>).

Procedure. Each participant completed 10 scanning runs of the main experiment that lasted just under 4 min each. To motivate fixation, participants were instructed to ignore the face stimuli flashing in the periphery and to perform a dimming task at fixation throughout. At the beginning of each run, a black fixation dot (0.1 degrees visual angle in diameter) was presented at the center of the screen on a gray background. Every 750 ms, the fixation dot changed its color for 500 ms from black to one of 13 colors (three shades of blue, green, yellow, orange, pink, three shades of purple, gray, white, and red). Participants were asked to press a button whether and only if the dot color changed to red.

In parallel to the ongoing fixation task, participants were presented with a face feature stimulus every 4 s. This amounted to 54 stimulus trials per run, 6 of which were baseline trials during which no stimulus was presented (apart from the ongoing fixation task). The remaining 48 stimulus trials were very similar to the recognition experiment, except for participants performing the fixation task instead of a recognition task.

Each run contained stimuli from two face photographs (i.e., two mouth stimuli and two pairs of eye stimuli) that randomly changed every two runs. Face feature stimuli (left eye/right eye/left mouth/right mouth with dummy coding for the latter when shown at the lower visual field position) and location (upper/lower visual field) were combined to result in eight trial types identical to the first recognition experiment and the corresponding four conditions (upper/lower visual field \times eye/mouth stimuli). Each of the eight trial types was repeated three times for stimuli from either face photograph, resulting in a total of 48 trials (plus six baseline trials) per run. The order of trials was pseudo-randomized, with the constraint that identical trials did not immediately follow each other.

At the beginning of each trial the stimulus image was briefly presented for 200 ms, overlaid by a dynamic noise mask that lasted until 450 ms after the stimulus image onset, just as in the recognition experiment. The mouth location was centered below fixation, on the vertical meridian at an eccentricity of 4 degrees visual angle. The left and right eye positions were centered 2.4 degrees above fixation and shifted laterally by 3.2 degrees, resulting in 4 degrees eccentricity. As in the recognition experiment, the exact stimulus location in each trial varied randomly with a scatter on the horizontal and vertical axes drawn from a Gaussian distribution (SD: 0.16 degrees visual angle; maximum scatter 2 SD).

Face localizer runs consisted of 20 alternating blocks of face and object images. Each block lasted just over 16 s (9 volumes at a TR of 1.785 s) and contained the 20 images of the respective set in randomized order, resulting in a run duration of just <5.5 min. Images were presented briefly for 400 ms with an interstimulus interval of 400 ms, whereas the black fixation dot was shown throughout. Participants were instructed to perform a 1-back task. In each block, a randomly determined image of the sequence was replaced by the preceding image of the sequence, resulting in a 1-back target. Participants were instructed to press a button whenever an image was identical to the immediately preceding one.

Image acquisition and preprocessing

All functional and structural scans were obtained with a TIM Trio 3T scanner (Siemens Medical Systems), using a 32-channel head coil. However, the front part of the head coil was removed for functional scans, leaving 20 effective channels (this way, restrictions of participants' field of view were minimized). Functional images for the main experiment were acquired with a gradient EPI sequence (2.3 mm isotropic resolution, matrix size 96×96 , 21 transverse slices per volume, acquired in

interleaved order and centered on the occipital and inferior temporal cortex; slice acquisition time 85 ms, TE 37 ms, TR 1.785 s). We acquired 129 volumes for each run of the main experiment and 185 volumes per localizer run. After five runs of the main experiment, B0 field maps were acquired to correct for geometric distortions in the functional images caused by heterogeneities in the B0 magnetic field (double-echo FLASH sequence with a short TE of 10 ms and a long TE of 12.46 ms, $3 \times 3 \times 2$ mm, 1 mm gap). Finally, two T1-weighted structural images were acquired for each participant. The first structural image was obtained with the front part of the head coil removed, using an MPRAGE sequence (1 mm isotropic resolution, 176 sagittal slices, matrix size 256×215 , TE 2.97 ms, TR 1900 ms). For the second structural image, the full 32-channel head coil was used with a 3D MDEFT sequence (Deichmann et al., 2004) (1 mm isotropic resolution, 176 sagittal partitions, matrix size 256×240 , TE 2.48 ms, TR 7.92 ms, TI 910 ms).

All image files were converted to NIFTI format and preprocessed using SPM 8 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>). The first five volumes for each run were discarded to allow for the T1 signal to reach steady state. The remaining functional images were mean bias corrected, realigned, unwarped (using voxel displacement maps generated from the fieldmaps) (Hutton et al., 2002), coregistered (with the respective anatomical MDEFT scan for each participant, using the MPRAGE scan as an intermediate step), and smoothed with a 5 mm Gaussian kernel. Images were preprocessed for each run independently. The anatomical MDEFT scan was used to reconstruct the cortical surface with FreeSurfer (<http://surfer.nmr.mgh.harvard.edu>).

Data analysis

To define ROIs, a GLM was specified for the localizer run of each participant. The model contained one regressor for blocks of face stimuli and one regressor for blocks of object stimuli (boxcar regressors convolved with a canonical hemodynamic response function). Additional regressors of no interest were modeled for the six motion parameters estimated during realignment. The parameters of the model were estimated and a contrast images and *t*-map calculated for the parameter contrast face-object stimuli for each participant.

The resulting *t*-maps were thresholded at $t > 2$ and projected onto the inflated cortical surface for each participant in FreeSurfer. Then contiguous patches of activation were identified and delineated at anatomical locations corresponding to those described by Weiner and Grill-Spector (2013): mid fusiform face area, posterior fusiform face area, and inferior IOG.

Because activations in IOG tended to be diffuse and the region could not clearly be delineated on functional grounds in 12 hemispheres (9 left and 3 right), the IOG ROI was instead defined anatomically on an individual basis. The FreeSurfer parcellation algorithm (Destrieux et al., 2010) was used to label the inferior occipital gyrus and sulcus on each hemisphere. The mean size (SEM) of these anatomical parcellations was 109.14 (5.09) and 91.14 (3.89) voxels for left and right hemispheres, respectively. Fusiform patches of activation could be detected in each of the 42 hemispheres. However, in 24 hemispheres, there was a single patch of fusiform activation that could not be separated into mid fusiform face area and posterior fusiform face area. Therefore fusiform patches of face-sensitive cortex were assigned a single fusiform face area (FFA) label for each hemisphere, resulting in a total of four ROIs per participant (left and right IOG as well as left and right FFA). All ROI labels were transformed into volume space and used as binary masks for the analyses described below.

Separate GLMs were run for each run and each participant. Each GLM contained regressors for each of the 8 trial types plus one regressor for baseline trials (boxcar regressors convolved with a canonical hemodynamic response function). Additional regressors of no interest were modeled for the six motion parameters estimated during realignment. The GLMs for each run and each participant were estimated and contrast images for each of the 8 trial types (per run) calculated. This resulted in separate contrast images and *t*-maps for each trial type, run, and participant. These *t*-maps were masked with the ROIs (see above), and the resulting patterns were vectorized (i.e., collapsed into a single row of data with entries corresponding to voxels in the original data space).

The aim of the decoding analysis was to decode stimulus identity from activation patterns (i.e., whether an eye or mouth stimulus was presented in a given trial) and to compare the accuracies of decoders across conditions (i.e., whether decoding accuracy varied for typical and reversed stimulus locations; for example pairs of stimuli, see Fig. 4D). Stimulus decoding was performed using custom code and the linear support vector machine implemented in the Bioinformatics toolbox for MATLAB (version R2013a, <http://www.mathworks.com>). Data from each condition were used for training and testing of separate classifiers to get condition-specific decoding accuracies. To avoid assigning a single classification label to stimuli shown at different locations, data within each condition were further subdivided according to visual hemifield. That is, data from left eye and left mouth stimuli were decoded separately as well as data from right eye and right mouth stimuli. For the typical condition, left and right mouth stimuli were dummy-coded (see above, Fig. 1C). A control analysis avoiding dummy coding by averaging the β weights of the respective regressors yielded the same pattern of results as the main analysis (for decoding as well as for amplitude results).

Classifiers were trained and tested for accuracy in a jack-knife procedure. In each iteration, the (condition and hemifield specific) data from all runs but one served as training data and the (condition and hemifield specific) data from the remaining run was used to test the prediction accuracy of the linear support vector machine. Accuracies were stored and averaged across iterations at the end of this procedure, and the whole procedure was applied to each ROI independently, yielding a single classification accuracy for each condition and ROI. Statistical analysis of the resulting accuracies was performed in MATLAB and PASW 20. Accuracies were compared against chance level by subtracting 0.5 and using one-sample t tests and against each other using a repeated-measures GLM with factors condition and hemifield, as well as *post hoc* t tests.

To obtain a fuller understanding of the dataset and to compare our data with previous studies (Kovács et al., 2008; Schwarzlose et al., 2008; Nichols et al., 2010), we ran additional analyses that tested the ability to decode feature identity separately for each location and the ability to decode stimulus location separately for each feature. That is, we aimed to decode eye- versus mouth-evoked patterns of activity, once for stimuli presented in the upper visual field and separately for stimuli presented in the lower visual field (see Fig. 5D). Similarly, we aimed to decode stimulus location (upper vs lower visual field), once for eye stimuli and separately for mouth stimuli (Fig. 5B). This analysis followed the same procedure as outlined for the main analysis above. The only difference was regarding which of the 8 trial types were paired for decoding and fed into the decoding routine (see above). To compare the resulting performance for “pure” location and “pure” feature decoding with the typical and reversed conditions from the main analysis, we determined average accuracies for each of these approaches in each ROI: that is, we collapsed accuracies across visual hemifields as well as stimuli (position decoding) or stimulus locations (feature decoding).

To test for differences in general amplitude levels, an additional mass-univariate analysis was performed. For this, a GLM was specified for each participant, including data from all 10 runs and regressors for each trial type and run as well as motion regressors for each run (see above). Additionally, the model contained regressors for the intercept of each run. To describe and compare general amplitude levels in either condition, four contrast images were derived for each participant. They corresponded to the contrast of trials from the typical or reversed conditions versus baseline trials, separately for stimuli shown in the left and right visual hemifield (dummy-coded as described above for mouths at the lower visual field position). For all contrast maps, the average values within each ROI were calculated for each participant. Then these averages were compared against 0 using one-sample t tests and against each other using a repeated-measures GLM with factors condition and hemifield, as well as *post hoc* t tests.

Finally, we tested hemisphere \times condition effects on decoding performance in a 2×4 repeated-measures GLM with factors hemisphere and type of decoding (including pure location and stimulus decoding; see above and Fig. 5), as well as a $2 \times 2 \times 2$ repeated-measures GLM with factors hemisphere, hemifield, and condition (see Fig. 4A). The same $2 \times$

2×2 repeated-measures GLM was performed with amplitude levels as dependent variable (see Fig. 4B).

To check fixation compliance during the fMRI experiment, we ran an additional univariate analysis testing retinotopic specificity of activations in early visual cortex on the single-participant level. For this analysis, we calculated contrast maps for stimuli shown in the lower versus upper visual field positions, as well as for stimuli shown in the contralateral versus ipsilateral upper visual field positions (for each hemisphere). These contrasts were derived using all stimuli shown at the respective positions (i.e., across both conditions). This yielded three t -maps per participant, each showing activations evoked by one of the stimulus positions in our design. We transformed these t -maps into binary activation maps and projected them onto the respective FreeSurfer reconstructed hemispheres (thresholded at $t > 2$ for three hemispheres with weaker activations and $t > 3$ for the rest). We then inspected the individual activation profiles in early visual cortex (EVC) (see Fig. 2G).

Hypothesis and predictions

Our main hypothesis was that eye- and mouth-preferring neural populations systematically differ with regard to their visual field preferences. Specifically, neural tuning for the eyes was expected to go along with a preference for the upper visual field and tuning for the mouth with a preference for the lower visual field (see Fig. 3B). We further hypothesized that these tuning properties are reflected at the level of BOLD (blood-oxygen-level-dependent) responses. That is, we expected a group of voxels with a response magnitude E preferring eyes over mouths, as well as the upper over the lower visual field. A second group of voxels (with a response magnitude M) was predicted to prefer mouths over eyes, as well as the lower over the upper visual field.

This immediately leads to a prediction of enhanced stimulus decoding for the typical compared with the reversed condition. We expected the maximum difference between E and M for the typical condition. In this condition, the stimulus will be preferred along both dimensions (retinal and face part tuning) by one group of voxels, but along neither dimension by the other. For example, an eye in the upper visual field will elicit maximum E , but minimum M . In contrast, a stimulus from the reversed condition would be expected to drive both responses to some extent. For example, an eye in the lower visual field would present a face part preferred by E , but at a retinal location preferred by M .

Predictions about net amplitude levels are less straightforward, because they involve the sum of E and M (and of all other populations in a ROI). Such predictions crucially depend on assumed integration rules for retinal and face part tuning. If this integration is additive, the expected amplitude difference between the typical and reversed conditions would be 0. That is, the sum of E and M would be identical in the typical and reversed conditions. Super-additive integration, on the other hand (e.g., multiplicative), would predict a net amplitude increase for the typical over the reversed condition (although simulations suggest that this amplitude increase would be small relative to the effect on decoding). Finally, subadditive integration following a principle of inverse effectiveness would predict a net amplitude decrease for the typical condition. That is, even if a stimulus preferred on both dimensions elicits the maximum response for a group of voxels (e.g., E for an eye in the upper visual field), this maximum response could be smaller than the sum of E and M for a stimulus preferred on only one dimension (e.g., an eye in the lower visual field).

It is not clear which integration rule retinotopic and face part tuning in BOLD responses of the human IOG might follow. In macaque, Freiwald et al. (2009) explicitly tested integration models for multidimensional tuning of face-sensitive neurons. They found that joint tuning curves were well predicted by a multiplicative model ($r = 0.89$), but only marginally worse by an additive model ($r = 0.88$). Integration rules for retinotopic and face part tuning specifically have not been tested but appear to be superadditive (compare Issa and DiCarlo, 2012, their Fig. 9). However, the integration of these dimensions in BOLD responses of the human IOG could still be additive or even subadditive (e.g., due to ceiling effects and as commonly observed for multisensory integration) (Stein and Stanford, 2008; Werner and Noppeney, 2011).

Together, net amplitude predictions can be zero, positive, or negative, depending on assumed integration rules. Crucially, the same is not true

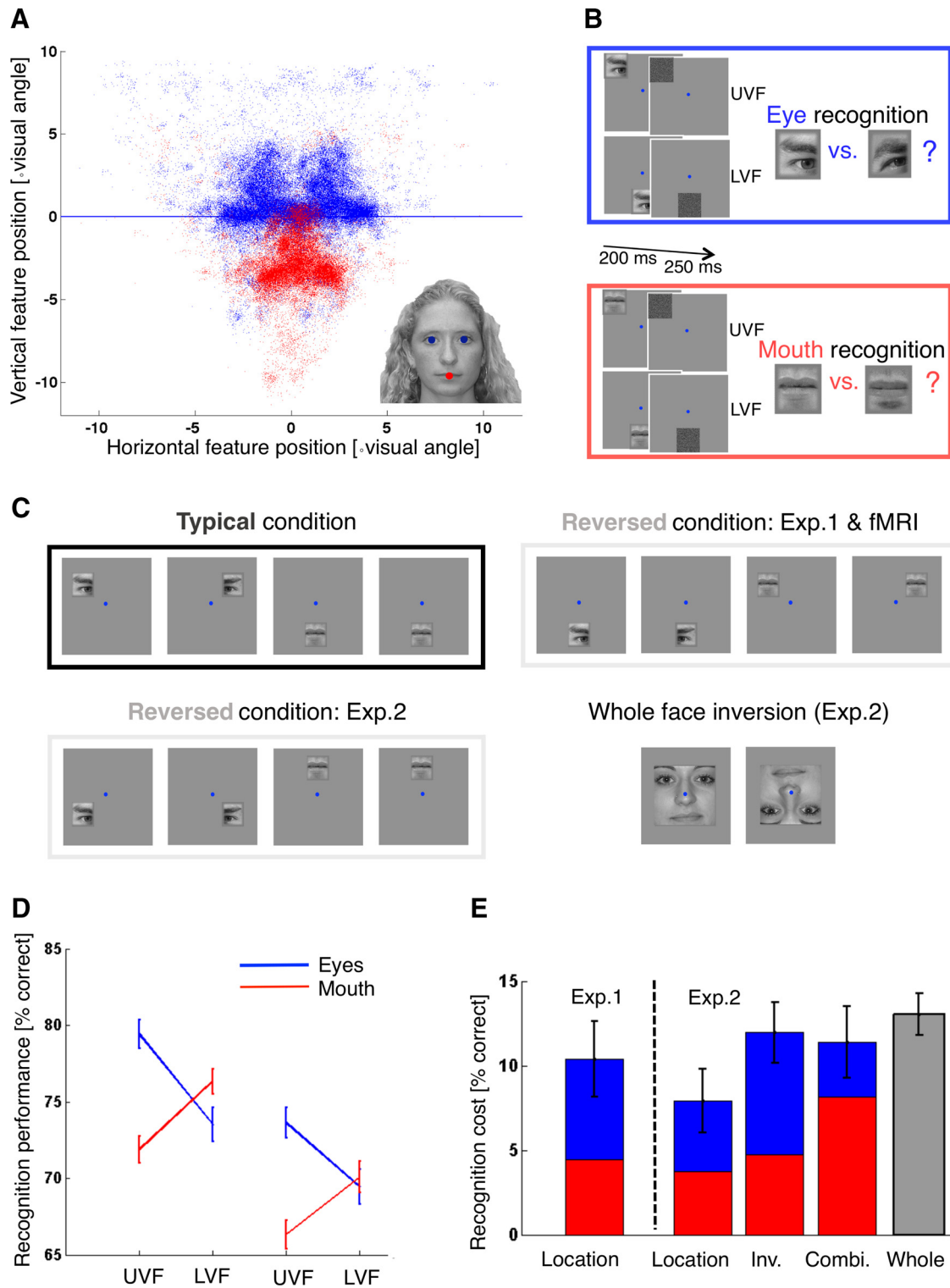


Figure 1. Recognition and gaze results. **A**, Visual field positions of eyes (blue) and mouth (red) across fixations. Participants freely viewed face images at conversational distance (80 cm) and size ($\sim 10 \times 14$ cm or $\sim 7 \times 10$ degrees visual angle; see Material and Methods). Inset, Example stimulus with the tracked eye and mouth features highlighted in blue and red. Blue line highlights the horizontal meridian. **B**, Recognition task. On each trial, participants briefly saw an image of an isolated eye or mouth, followed by a dynamic noise mask. Participants fixated on a dot at the center of the screen (shown in blue), and the images could appear at visual field positions typical for an eye (UVF, Upper visual field) or the mouth (LVF, Lower visual field). The task was to indicate which of two candidate eye or mouth images was shown. Fixation compliance was ensured using an eye tracker (Fig. 2). **C**, Stimulus positions. The layout of stimulus position in the typical condition was common across experiments (top left). Left and right eye, as well as mouth images, were flashed at the respective typical retinal positions (which were iso-eccentric from the fixation point). The typical mouth trial is shown twice to indicate that the sum of typical mouth trials was the same as that for typical left and right eye trials. Trials from the reversed condition in the first recognition and fMRI experiments showed left and right eye images at the mouth position and mouth images at the left and right eye positions (top right). Reversed stimulus locations in the second recognition experiment were closer to those in an inverted face (bottom left). The second recognition experiment also tested the effect of whole face inversion, showing upright and inverted faces (bottom right). Additionally, the second recognition experiment showed inverted features at the typical and reversed locations shown here (inverted features not shown). **D**, Recognition performance. Data points and error bars indicate mean recognition performance \pm SEM as percentage correct. Performance for eye and mouth recognition is shown in blue (*Figure legend continues.*)

for decoding. Our main hypothesis predicts a decoding advantage for the typical over the reversed condition, regardless of assumed integration rule.

Results

Face features appear at typical visual field positions

We tested and quantified the retinotopic bias for eye and mouth features from eye tracking data of 14 healthy adults freely viewing 24 face images of approximately natural size and at conversational distance (see Materials and Methods; Fig. 1A). Across observers, face stimuli, and fixation events, eyes appeared in the upper visual field $76 \pm 4\%$ of the time ($t_{(13)} = 6.78, p = 1.3 \times 10^{-5}$) and mouths in the lower visual field $94 \pm 2\%$ of the time ($t_{(13)} = -26.50, p = 1.1 \times 10^{-12}$); the mean height of eyes and mouths was 0.57 ± 0.10 and 3.31 ± 0.13 degrees visual angle above and below fixations, respectively ($t_{(13)} = 5.48, p = 1.1 \times 10^{-4}$ and $t_{(13)} = -26.15, p = 1.3 \times 10^{-12}$).

Recognition of face features is better at typical visual field positions

Next, we tested whether perceptual sensitivity for eyes and mouths was tuned to these typical visual field positions. Eighteen participants saw briefly presented eye or mouth images (see Materials and Methods), appearing at typical or reversed upper and lower visual field locations while fixation was ensured using an eyetracker. Participants performed a two-alternative forced choice recognition task, deciding on each trial which of two different eye or mouth images they saw (see Materials and Methods; Fig. 1B). There was a significant feature \times location interaction for recognition performance ($F_{(1,17)} = 21.87, p = 2.1 \times 10^{-4}$). Specifically, *post hoc t* tests revealed that participants individuated eyes significantly better in the upper than lower visual field ($79.42 \pm 2.25\%$ vs $73.51 \pm 2.23\%$ correct; $t_{(17)} = 3.34, p = 0.004$), whereas the reverse was true for individuating mouth stimuli ($71.86 \pm 1.77\%$ vs $76.34 \pm 1.62\%$ correct; $t_{(17)} = -3.40, p = 0.003$; Fig. 1D, left). There also was a small recognition advantage for eye over mouth stimuli ($76.47 \pm 2.1\%$ vs $74.10 \pm 1.56\%$ correct; $F_{(1,17)} = 4.95, p = 0.04$), but no main effect of stimulus position ($F_{(1,17)} = 0.43, p = 0.52$). This effect was replicated in a larger sample of 36 participants during a second recognition experiment, in which the reversed visual field positions more closely resembled those in inverted faces (see Materials and Methods). There was a significant feature \times location interaction for recognition performance ($F_{(1,35)} = 18.07, p = 1.5 \times 10^{-4}$), with a trend for better eye-individuation in the upper visual field ($73.61 \pm 1.59\%$ vs $69.44 \pm 1.81\%$ correct; $t_{(35)} = 1.92, p = 0.09$, not significant) and the opposite effect for mouth individuation ($66.32 \pm 1.45\%$ vs $70.08 \pm 1.58\%$ correct; $t_{(35)} = -3.27, p = 0.009$; Fig. 1D, right). There also was a recognition advantage for eye over mouth stimuli ($71.53 \pm 1.53\%$ vs $68.20 \pm 1.32\%$ correct; $F_{(1,35)} = 7.44, p = 0.01$), but no main effect of stimulus position

($F_{(1,35)} = 0.30, p = 0.86$). Across conditions, recognition performance in Experiment 2 was worse than in Experiment 1, possibly due to the somewhat greater stimulus eccentricity.

In Experiment 2, we also compared the recognition cost of feature-based manipulations with that of whole face inversion. Specifically, we investigated the recognition cost of feature inversion, reversed feature locations, and the combination of both. Recognition performance for upright and inverted whole faces was $87.67 \pm 1.41\%$ and $74.65 \pm 1.73\%$, respectively, resulting in a recognition cost of $13.02 \pm 1.24\%$ for whole faces ($t_{(35)} = 10.52, p = 2.21 \times 10^{-12}$; Fig. 1E). The summed recognition cost of reversed feature locations (across eye and mouth effects) was $7.93 \pm 1.86\%$ (Experiment 1: $10.39 \pm 2.22\%$), that of feature inversion $11.98 \pm 1.78\%$, and that of combining inversion and reversed feature locations $11.40 \pm 2.08\%$. Comparing the recognition cost of feature manipulations yielded no statistically significant main effect of feature (eye/mouth; $F_{(1,35)} = 0.05, p = 0.82$) or manipulation (reversed location/inversion/combined inversion and location reversal; $F_{(1,35)} = 1.36, p = 0.26$). However, there was a significant feature \times manipulation interaction ($F_{(1,35)} = 5.99, p = 0.004$). *Post hoc t* tests showed that for mouth stimuli the recognition cost of the combined manipulation was higher than for either location reversal ($t_{(35)} = 2.63, p = 0.01$) or feature inversion ($t_{(35)} = 2.04, p = 0.049$) on its own. There was no significant difference between the effects of location reversal and feature inversion ($t_{(35)} = 0.77, p = 0.45$). For eye stimuli, isolated feature inversion resulted in a larger recognition cost than the combined manipulation ($t_{(35)} = 2.51, p = 0.01$), and there was a nonsignificant trend for larger inversion than location effects ($t_{(35)} = 1.83, p = 0.08$). There was no significant difference between the recognition cost of location reversal and the combined manipulation ($t_{(35)} = 0.59, p = 0.56$).

Fixation compliance was good in both recognition experiments, with an average median absolute deviation from fixation of $0.52 (\pm 0.06 \text{ SEM}; \text{Experiment 1})$ and $0.43 (\pm 0.05 \text{ SEM}; \text{Experiment 2})$ degrees visual angle on the horizontal axis and $0.77 (\pm 0.13 \text{ SEM}; \text{Experiment 1})$ and $0.56 (\pm 0.08 \text{ SEM}; \text{Experiment 2})$ degrees visual angle on the vertical axes (Fig. 2).

Stimulus discrimination in right IOG is better for typical visual field positions

We hypothesized that the spatial heterogeneity in perceptual sensitivity is reflected in a corresponding link between feature and retinotopic tuning in the underlying neural populations. We expected such neural tuning to occur most likely in the IOG, containing the occipital face area (OFA) (Pitcher et al., 2011). The OFA is sensitive to face parts (Pitcher et al., 2007; Liu et al., 2010; Nichols et al., 2010; Orlov et al., 2010; Henriksson et al., 2015) as well as stimulus location (Schwarzlose et al., 2008; Kay et al., 2015). Furthermore, the OFA is thought to be the homolog of the macaque posterior lateral face patch (Tsao et al., 2008; Issa and DiCarlo, 2012), which contains cells that show a preference for stimuli depicting the contralateral eye, as well as for the contralateral upper visual field quadrant (at least for early response components) (Issa and DiCarlo, 2012).

We reasoned that a link between neural spatial and feature preferences would affect pattern separability at the voxel level. If feature and location preferences are correlated, response patterns to different features should be better separable when the features are presented at typical rather than reversed locations and vice versa (Fig. 3; see Hypothesis and predictions). In an fMRI experiment, 21 participants passively viewed brief flashes of eye and mouth stimuli at either typical or reversed visual field locations

←
(Figure legend continued.) and red and shown for the upper (UVF) and lower (LVF) visual field. Left, Data from Experiment 1. Right, Data from Experiment 2. Overall, recognition performance in Experiment 2 was worse than Experiment 1 (possibly due to greater stimulus eccentricity), but the hypothesized interaction effect was replicated. **E**, Effect sizes. Bars and error bars indicate the mean recognition cost (\pm SEM) of different feature manipulations (relative to the upright, typical location version of a given feature): location reversal (Location), feature inversion (Inv.), and the combination of feature inversion and reversal (Combi.). Blue and red bar segments represent recognition cost for eye and mouth features, respectively. Rightmost bar represents mean recognition cost (\pm SEM) for the inversion of whole face images (face inversion effect size).

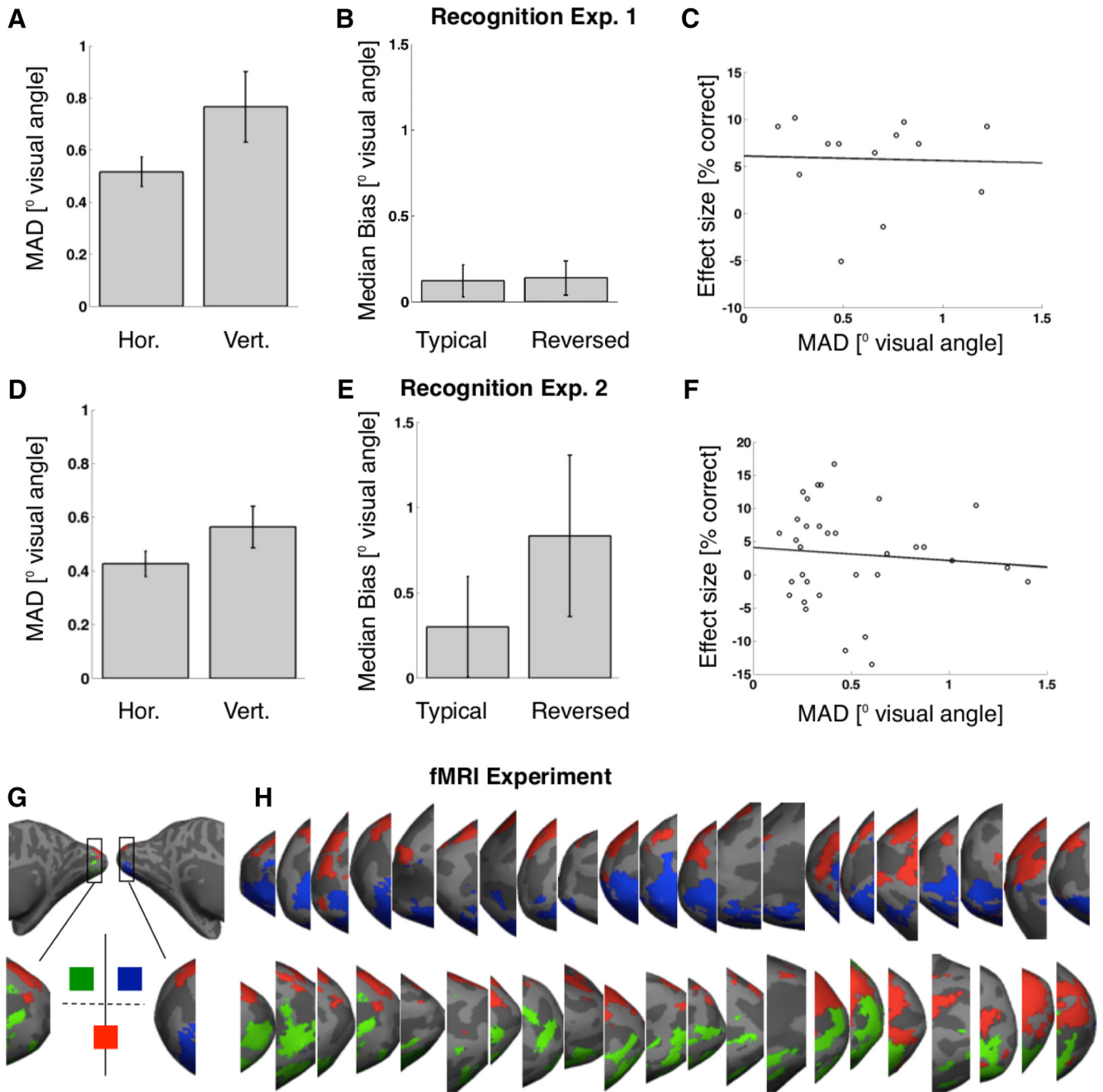


Figure 2. Fixation compliance. Top (A–C) and bottom (D–F) rows represent data from recognition Experiments 1 and 2, respectively. Right panels (G–J) represent the results of a gaze control analysis for the fMRI experiment. **A, C**, Median absolute deviation (MAD) from fixation during stimulus presentation (in degrees visual angle). Left and right bars represent deviation along the horizontal and vertical axes, respectively. Generally, fixation compliance was good, with average median absolute deviations from fixation of $0.52 (\pm 0.06 \text{ SEM}; \text{Experiment 1})$ and $0.43 (\pm 0.05 \text{ SEM}; \text{Experiment 2})$ degrees visual angle on the horizontal axis and $0.77 (\pm 0.13 \text{ SEM}; \text{Experiment 1})$ and $0.56 (\pm 0.08 \text{ SEM}; \text{Experiment 2})$ degrees visual angle on the vertical axes. **B, E**, Median bias of gaze direction toward stimuli (along the vertical axis and in degrees visual angle). Left and right bars represent bias during presentation of stimuli in typical and reversed locations, respectively. Deviation from fixation was not significantly biased toward stimuli in either typical (Experiment 1: average bias $0.12 (\pm 0.09 \text{ SEM})$ degrees visual angle, $t_{(12)} = 1.30, p = 0.22$; Experiment 2: $0.30 (\pm 0.29 \text{ SEM})$ degrees visual angle, $t_{(32)} = 1.02, p = 0.32$) or reversed trials (Experiment 1: $0.14 (\pm 0.10 \text{ SEM})$ degrees visual angle, $t_{(12)} = 1.39, p = 0.19$); although there was a nonsignificant trend in Experiment 2: $0.83 (\pm 0.47 \text{ SEM})$ degrees visual angle, $t_{(32)} = 1.76, p = 0.09$). There was no significant difference in bias between conditions either (Experiment 1: $t_{(12)} = -0.89, p = 0.39$; Experiment 2: $t_{(31)} = -1.74, p = 0.09$). **C, F**, Scatter plot of individual effect size versus individual deviation from fixation. Effect size is defined as the recognition advantage for stimuli presented at typical versus reversed locations (in percentage correct answers). Deviation from fixation is the median absolute deviation in degrees visual angle (averaged across the horizontal and vertical axes). Each circle represents data from one participant. A line indicates the least-squares fit to the data. Individual differences in the propensity for eye movements did not predict individual differences in the size of the hypothesized effect in Experiment 1 ($r_{(11)} = -0.04, p = 0.91$) or 2 ($r_{(31)} = -0.09, p = 0.62$). **G**, Retinotopic control analysis testing fixation compliance during the fMRI experiment. We projected binary activation maps onto the inflated cortical surfaces, each map corresponding to one of the three stimulus locations in our design (compare Materials and Methods). Green, blue, and red activation maps represent the left, right upper, and the lower visual field position, respectively, as shown. We then investigated the activation profile in early visual cortex for each participant (i.e., medial, posterior occipital, surrounding the calcarine sulcus). **H**, Activation profiles in left (top row) and right (bottom row) EVC of each participant. As expected under fixation compliance, stimuli shown at the lower visual field position evoked bilateral dorsal activations, and stimuli shown at the upper visual field positions activated contralateral ventral EVC.

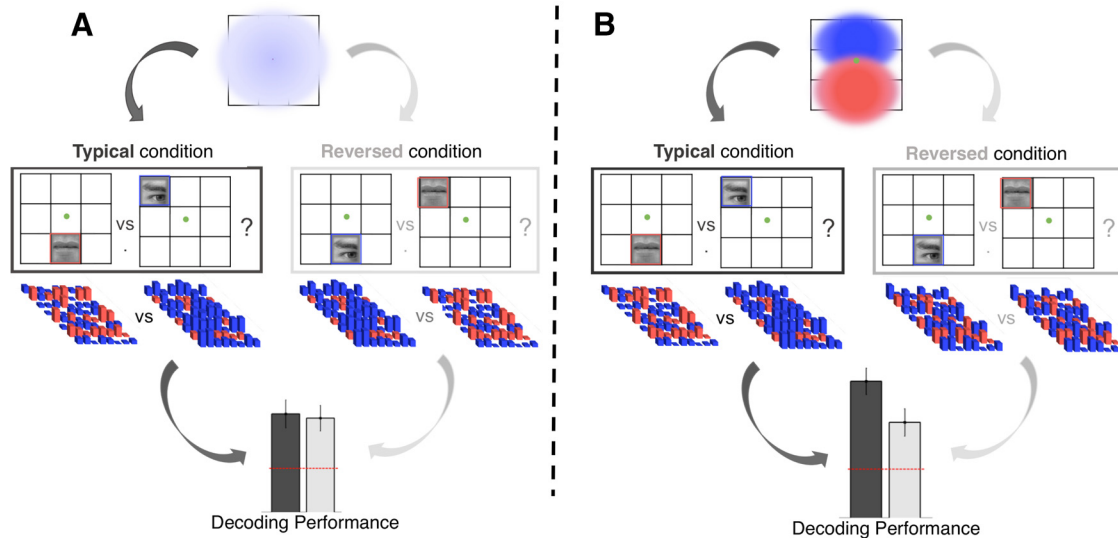


Figure 3. Hypotheses and predictions for neuroimaging experiment. **A**, Null hypothesis. Eye- and mouth-prefering neural units do not differ in their spatial preferences. Population receptive fields of either feature type cover the entire central visual field (top). Therefore, patterns of voxel activations evoked by eye and mouth stimuli will differ to the same degree, regardless of whether stimuli are shown at typical (dark gray) or reversed (light gray) visual field locations (middle). Classification performance based on the separability of these patterns will not be affected by condition (bottom). **B**, Alternative hypothesis. Population receptive fields of eye-prefering (blue) and mouth-prefering (red) units show a preference for the upper and lower visual field, respectively (top). Therefore, patterns of voxel activations evoked by eye and mouth stimuli will differ more when stimuli are shown at typical (dark gray) rather than reversed (light gray) locations and both their differential feature and spatial preferences are met (middle). Classification performance based on the separability of these patterns will be better in the typical compared with the reversed condition (bottom). For details, see Hypothesis and predictions.

Table 1. Classification performance by condition and hemifield for each ROI^a

	Typical condition		Reversed condition		Main effect hemifield	Main effect condition	
	Left hemifield	Right hemifield	Left hemifield	Right hemifield	Right-left	Typical-reversed	Interaction
Left IOG	58.33% (3.09%) $t_{(20)} = 2.69, p = 0.01$	60.71% (2.83%) $t_{(20)} = 3.79, p = 0.001$	58.10% (2.08%) $t_{(20)} = 3.88, p < 0.001$	63.33% (2.75%) $t_{(20)} = 4.86, p < 0.001$	3.81% (2.93%) $t_{(20)} = 1.30, NS$	-1.19% (2.08%) $t_{(20)} = -0.57, NS$	$F_{(1,20)} = 0.50, NS$
Right IOG	59.52% (2.41%) $t_{(20)} = 3.95, p < 0.001$	56.67% (1.7%) $t_{(20)} = 3.92, p < 0.001$	55.00% (2.46%) $t_{(20)} = 2.03, NS$	53.81% (2.01%) $t_{(20)} = 1.90, NS$	-2.02% (2.34%) $t_{(20)} = -0.86, NS$	3.69% (1.68%) $t_{(20)} = 2.20, p = 0.04$	$F_{(1,20)} = 0.22, NS$
Left FFA	55.48% (1.65%) $t_{(20)} = 3.32, p = 0.003$	53.81% (2.36%) $t_{(20)} = 1.61, NS$	53.10% (1.78%) $t_{(20)} = 1.74, NS$	53.57% (2.44%) $t_{(20)} = 1.46, NS$	-0.6% (2.52%) $t_{(20)} = -0.24, NS$	1.31% (2.00%) $t_{(20)} = 0.65, NS$	$F_{(1,20)} = 0.41, NS$
Right FFA	56.19% (1.89%) $t_{(20)} = 3.28, p = 0.004$	52.38% (2.55%) $t_{(20)} = 0.93, NS$	49.76% (1.40%) $t_{(20)} = -0.17, NS$	52.38% (2.03%) $t_{(20)} = 1.17, NS$	-0.6% (1.82%) $t_{(20)} = -0.33, NS$	3.21% (1.67%) $t_{(20)} = 1.92, NS$	$F_{(1,20)} = 7.52, p = 0.01$

^aCells contain the mean (SEM) across participants for decoding performance of a linear support vector machine classifying stimulus-evoked patterns of BOLD (blood-oxygen-level-dependent) activation in an ROI. t and p values in the first four columns correspond to one-sample t test of decoding performance against chance level (50%). The fifth and sixth columns give the mean (SEM) for a decoding advantage for stimuli shown in the right versus left visual hemifield or typical versus reversed conditions, respectively. t and p values in this column correspond to a paired-sample t test comparing decoding performance between hemifields or conditions. The right-most column gives F values of a repeated-measures GLM testing for an interaction between hemifield and condition. Compare Figure 4A.

while solving an incidental fixation task (see Materials and Methods). We trained linear support vector machines to distinguish patterns of BOLD signals evoked by eye or mouth stimuli in the IOG and the FFA of each hemisphere, which were determined independently for each participant. BOLD signal patterns were separable by stimulus type significantly above chance in all ROIs (Table 1; Fig. 4A). Crucially, pattern separability in right IOG was significantly better for typical compared with reversed stimulus locations ($t_{(20)} = 2.20, p = 0.04$), and there was a similar trend for the right FFA ($t_{(20)} = 1.92, p = 0.07$). Furthermore, there was a significant hemifield \times condition interaction in right FFA ($F_{(1,20)} = 7.52, p = 0.01$), with a strong decoding advantage for typical compared with reversed stimulus locations in the contralateral left ($t_{(20)} = 3.39, p = 0.002$) but not the ipsilateral right ($t_{(20)} < 0.01, p > 0.99$) visual hemifield.

Univariate analyses confirmed that every ROI responded to the eye or mouth stimuli (with the exception of responses to ipsilateral stimuli from the typical condition in left IOG; $t_{(20)} = 1.68, p = 0.11$; all other: $t > 3, p < 0.01$). Further, there were significant main effects of hemifield in left IOG and left FFA, with higher response for stimuli presented in the contralateral right

visual hemifield in both areas ($F_{(1,20)} = 13.20, p = 0.002$ and $F_{(1,20)} = 7.27, p = 0.01$, respectively). Finally, there was a trend toward higher response amplitudes for the reversed compared with the typical condition in all ROIs, and this difference was significant in the right FFA ($F_{(1,20)} = 4.41, p = 0.048$; Table 2; Fig. 4B). We further tested potential hemispheric differences using a $2 \times 2 \times 2$ repeated-measures GLM with factors hemisphere, hemifield, and condition (compare Materials and Methods; Fig. 4B). This analysis confirmed a hemisphere \times hemifield interaction in IOG ($F_{(1,20)} = 48.41, p < 0.001$) and FFA ($F_{(1,20)} = 43.23, p < 0.001$), with hemifield significantly modulating response amplitude in the left but not the right hemisphere. It further showed a significant main effect of hemisphere for IOG, with higher amplitudes in the right hemisphere ($F_{(1,20)} = 10.52, p = 0.004$; Fig. 4B), but no other significant main or interaction effects.

To obtain a more complete understanding of the data, we additionally probed the ability to decode “pure” location (stimuli in upper vs lower visual field) for constant features (i.e., decoded separately for eyes and mouths) and vice versa “pure” feature identity for constant stimulus locations (see Materials and Methods; Fig. 5). Location could be decoded significantly above

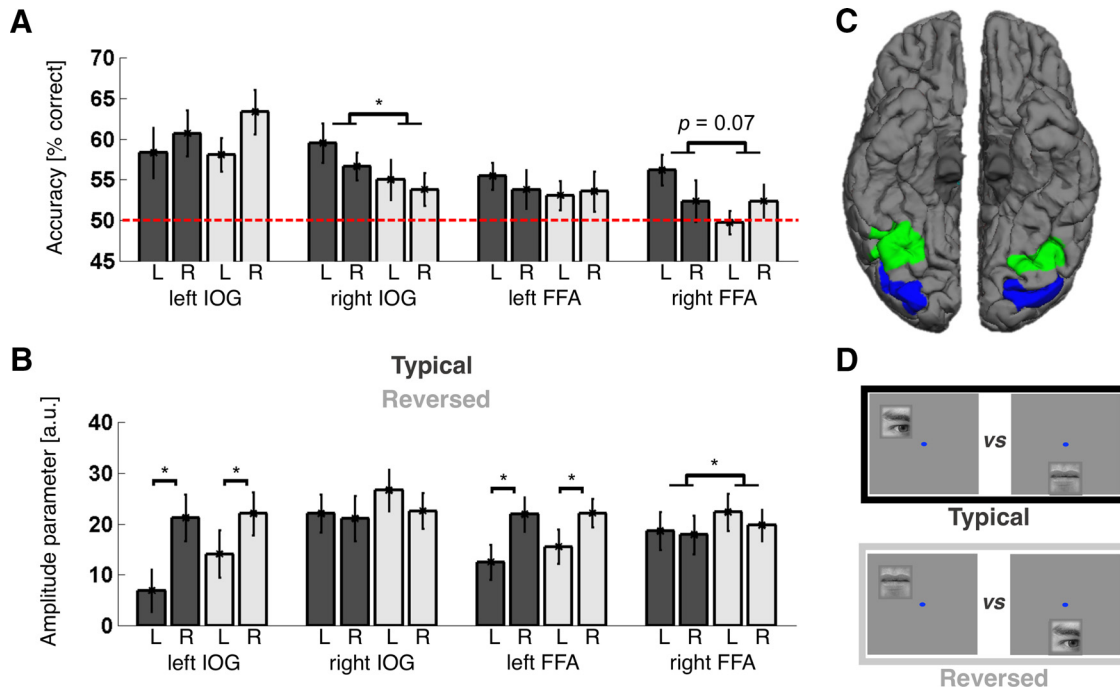


Figure 4. Main neuroimaging results. **A**, Classification performance by condition and visual hemifield for each ROI. Bar plots and error bars indicate mean classification performance \pm SEM for separating eye- and mouth-evoked patterns in the typical (dark gray) and reversed (light gray) conditions (for example stimuli, see **D**). Decoding performance for stimuli shown in the left and right visual hemifield is shown side by side for each ROI and condition (labeled L and R, respectively). The hemifield was dummy-coded for mouth stimuli, which were shown on the midline (see Materials and Methods). Dashed red line indicates chance level. ROIs are labeled on the x-axis. Classification performance was significantly better for the typical than reversed condition in right IOG ($p = 0.04$), and there was a similar trend in right FFA ($p = 0.07$). Furthermore, there was a significant hemifield \times condition interaction in right FFA ($p = 0.01$), with a strong decoding advantage for typical compared with reversed stimulus locations in the contralateral left ($p = 0.002$), but not the ipsilateral right ($p > 0.99$) visual hemifield (compare Table 1). $*p < 0.05$. **B**, Response amplitude by condition and ROI. Values on the y-axis indicate response amplitudes elicited by face feature stimuli versus baseline trials in arbitrary units; bars and error bars indicate the mean and SEM across participants. Bars in dark and light gray represent the typical and reversed conditions, respectively, L and R indicate amplitudes for stimuli shown in the left and right visual hemifield, respectively. ROIs are labeled on the x-axis. Response amplitudes were significantly greater than 0 in all regions; left IOG and left FFA responded significantly stronger to stimuli in the contralateral right visual hemifield, and there was an overall trend for slightly stronger responses to stimuli shown at reversed compared with typical conditions (which was significant in right FFA, $p = 0.048$; compare Table 2). $*p < 0.05$. **C**, ROIs. Image represents the inferior view of both hemispheres for an example participant. Blue represents the IOG. Green represents the FFA. Surface reconstruction was done using FreeSurfer (compare Materials and Methods). **D**, Conditions. Top, Example pair of stimuli from the typical condition, showing eye and mouth features at typical locations relative to a fixation dot (shown in blue). Bottom, Position of the same features in the reversed condition. Classification aimed at separating patterns of BOLD signals evoked by one or the other member of such pairs.

Table 2. Response amplitude by condition and hemifield for each ROI^a

	Typical condition		Reversed condition		Main effect hemifield	Main effect condition	Interaction
	Left hemifield	Right hemifield	Left hemifield	Right hemifield			
Left IOG	6.93 (4.13) $t_{(20)} = 1.68$, NS	21.22 (4.55) $t_{(20)} = 4.66$, $p < 0.001$	14.1 (4.68) $t_{(20)} = 3.01$, $p = 0.006$	22.01 (4.24) $t_{(20)} = 5.2$, $p < 0.001$	22.19 (6.11) $t_{(20)} = 3.63$, $p = 0.002$	7.96 (4.14) $t_{(20)} = 1.92$, NS	$F_{(1,20)} = 3.29$, NS
Right IOG	22.07 (3.77) $t_{(20)} = 5.86$, $p < 0.001$	21.08 (4.47) $t_{(20)} = 4.72$, $p < 0.001$	26.63 (4.06) $t_{(20)} = 6.56$, $p < 0.001$	22.57 (3.45) $t_{(20)} = 6.54$, $p < 0.001$	-5.06 (7.22) $t_{(20)} = -0.7$, NS	6.04 (3.61) $t_{(20)} = 1.67$, NS	$F_{(1,20)} = 0.55$, NS
Left FFA	12.53 (3.44) $t_{(20)} = 3.65$, $p = 0.002$	21.88 (3.42) $t_{(20)} = 6.41$, $p < 0.001$	15.53 (3.4) $t_{(20)} = 4.57$, $p < 0.001$	22.12 (2.78) $t_{(20)} = 7.96$, $p < 0.001$	15.93 (5.91) $t_{(20)} = 2.7$, $p = 0.01$	3.24 (3.39) $t_{(20)} = 0.96$, NS	$F_{(1,20)} = 0.73$, NS
Right FFA	18.63 (3.69) $t_{(20)} = 5.05$, $p < 0.001$	17.86 (3.79) $t_{(20)} = 4.71$, $p < 0.001$	22.33 (3.64) $t_{(20)} = 6.14$, $p < 0.001$	19.76 (3.1) $t_{(20)} = 6.38$, $p < 0.001$	-3.34 (5.28) $t_{(20)} = -0.63$, NS	5.6 (2.67) $t_{(20)} = 2.1$, $p = 0.048$	$F_{(1,20)} = 0.36$, NS

^aCells contain the mean (SEM) across participants for BOLD (blood-oxygen-level-dependent) amplitudes elicited by eye or mouth stimuli in an ROI (compared with baseline trials; arbitrary units; see Materials and Methods). t and p values in the left four columns correspond to one-sample t test of response amplitude against 0. The fifth and sixth columns give the mean (SEM) for the difference in response amplitudes for stimuli presented in the right-left visual hemifield and for the reversed-typical conditions, respectively (for details of conditions, see Materials and Methods). t and p values in these columns correspond to a paired-sample t test comparing response amplitudes between hemifields or conditions. The right-most column gives F values of a repeated-measures GLM testing for an interaction between hemifield and condition. Compare Figure 4B.

chance level in left IOG ($60.06 \pm 1.46\%$ correct; $t_{(20)} = 6.90$, $p < 0.001$), right IOG ($55.24 \pm 0.99\%$ correct; $t_{(20)} = 5.27$, $p < 0.001$), and left FFA ($53.87 \pm 0.99\%$ correct; $t_{(20)} = 3.91$, $p < 0.001$), but not right FFA ($51.90 \pm 1.51\%$ correct; $t_{(20)} = 1.26$, $p = 0.22$, not significant). Feature identity (eye vs mouth) could not be decoded from any of the ROIs (all: $t_{(20)} < 0.80$, all $p > 0.43$). In right IOG, where decoding performance was significantly better for the typical versus reversed condition, decoding stimuli at typical locations yielded significantly higher accuracy than pure location decoding ($t_{(20)} = 2.16$, $p = 0.04$), but there

was no significant difference between pure location decoding and the reversed condition ($t_{(20)} = 0.52$, $p = 0.61$). In right FFA, where we observed a similar trend toward better decoding for typical stimulus locations, decoding performance was significantly better than chance only for the typical condition ($t_{(20)} = 2.29$, $p = 0.03$). However, the observed advantage for the typical condition over pure location decoding was not significant ($t_{(20)} = 1.10$, $p = 0.29$).

Finally, we tested hemisphere \times condition effects on decoding performance. First we tested a 2×4 repeated-measures general

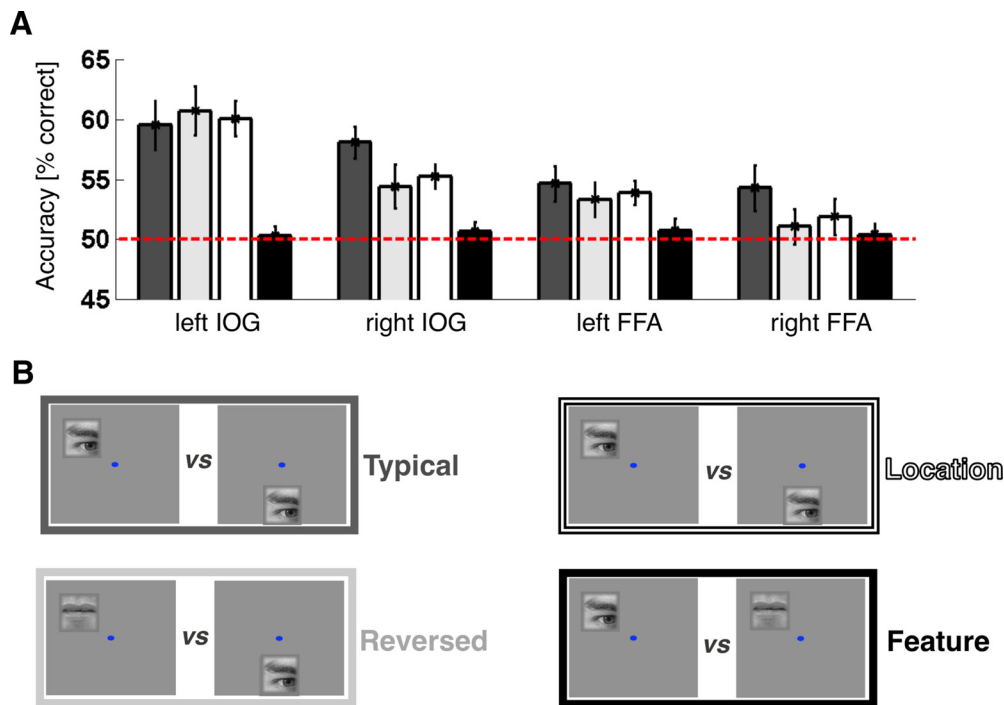


Figure 5. Additional neuroimaging results. **A**, Classification performance by decoding approach and ROI. Bar plots and error bars indicate mean classification performance \pm SEM for separating patterns evoked by different trial types. **B**, Example pairs of decoded stimuli (for more details, see and Methods, Results). Classification performance for typical location-feature pairings, reversed pairings, “pure” location (constant features) and “pure” feature information (constant location), shown in dark gray, light gray, white, and black, respectively. Dashed red line indicates chance level. ROIs are labeled on the x-axis.

linear model across the four types of decoding shown in Figure 5. For FFA there was no significant main effect of hemisphere ($F_{(1,20)} = 1.78, p = 0.20$) and no interaction effect between hemisphere and condition ($F_{(3,60)} = 0.47, p = 0.70$). However, there was a significant main effect of better overall decoding in left compared with right IOG ($F_{(1,20)} = 8.12, p = 0.01$). There also was a significant interaction between hemisphere and condition in IOG ($F_{(3,60)} = 4.56, p = 0.006$), pointing to the decoding advantage for the typical condition over pure location decoding and the reversed condition in right but not left IOG (Fig. 5). However, there was no significant hemisphere \times condition interaction for decoding accuracy in IOG ($F_{(1,20)} = 3.88, p = 0.06$) or FFA ($F_{(1,20)} = 3.59, p = 0.07$) when tested in the context of a $2 \times 2 \times 2$ repeated-measures GLM with factors hemisphere, hemifield, and condition (as for amplitudes, see above and Fig. 4B). This analysis showed a significant main effect of hemisphere in IOG (with better decoding in the left hemisphere; $F_{(1,20)} = 6.62, p = 0.02$), but no other significant effects.

To check gaze stability in the scanner, we ran a retinotopic control analysis on the single participant level. This analysis confirmed the expected, retinotopically specific activations in EVC (Fig. 2G,H). Stimuli shown at the lower visual field position activated bilateral dorsal EVC and stimuli shown at the upper visual field positions activated contralateral ventral EVC. This pattern held for every participant, confirming good fixation compliance (Fig. 2H).

Discussion

We observed a canonical retinotopic bias for features of faces presented at conversational distance. Furthermore, observers were significantly better at individuating eye images at upper compared with lower visual field locations, whereas the opposite was true for mouth images. Similarly, eye- and mouth-evoked patterns of BOLD signals in right IOG were decodable with

higher accuracy when stimuli were presented at typical, rather than reversed, visual field locations. We conclude that face perception is tuned to typical retinotopic feature positions and that the underlying neural mechanism likely is a corresponding match of spatial and feature preferences in right IOG. Our results are consistent with and extend previous findings. Early responses in the posterior lateral face patch of macaque show a preference for the contralateral eye region of a face (Issa and DiCarlo, 2012) as well as for the contralateral upper visual field quadrant. Moreover, half-faces elicit more distinct patterns of BOLD activity in right FFA when they are presented in the (typical) contralateral hemifield (Chan et al., 2010). At a finer scale, patches within the occipital face area that respond to different face features appear to be laid out in an arrangement resembling a face (Henriksson et al., 2015).

The present results extend these findings by showing differential spatial tuning for the perceptual and neural processing of eyes and mouth in humans. While previous results showed eye-preferring neurons in macaque are tuned to the upper visual field (Issa and DiCarlo, 2012), they did not address whether this effect had any perceptual relevance. Further, it was an open question whether such a match extends to face parts other than the eyes and to human observers. Our data suggest positive answers to all of these questions. Previous studies have established the prime importance of eyes and mouth for face processing (Smith et al., 2004, 2009) and the importance of their relative spatial positioning (Tanaka and Farah, 1993; Liu et al., 2010). Here we show that face perception also depends on absolute, retinotopic feature placement, with better performance for feature-specific canonical locations.

Importantly, the interaction nature of our main hypothesis controlled for main effects of stimulus or location. It was the

combination of features and locations that determined performance. The balanced nature of our design further ensured that stimuli and locations were unpredictable on any given trial. A disadvantage of our design is that we tested only a limited range of locations. Future studies should sample the visual field more densely and test potential links with individual gaze behavior (compare Peterson and Eckstein, 2013). Also, future studies should test how retinotopic feature locations interact with their relative positioning (e.g., to a face outline) (Issa and DiCarlo, 2012).

We hypothesized that the perceptual effect we observed is based on correlated neural preferences for facial features and locations. This led us to predict that stimulus-evoked patterns of BOLD activity would be more distinct for facial features presented at typical, compared with reversed, retinotopic locations. Indeed, decoding accuracy was significantly better for the typical condition in right IOG, and there was a similar trend in right FFA. This might point to a homology between human IOG and the posterior lateral face patch in macaque (Issa and DiCarlo, 2012). However, Issa and DiCarlo (2012) and Freiwald et al. (2009) suggest a near total dominance of the eye region in macaque face processing. Our recognition experiments revealed clear retinal preferences for mouth (as well as eye) perception in humans, pointing to a potential species difference.

Furthermore, our findings are consistent with models of object recognition that propose that representations of object identity, location, and structure depend on an (acquired) combination of neural tuning properties (Edelman and Intrator, 2000; Ullman, 2007; Kravitz et al., 2008). Combining feature and location preferences in this way seems efficient: predictable retinotopic feature locations would render a complete sampling mosaic of face features across the visual field excessive. Future studies could test whether similar contingencies hold for other stimuli for which observers have canonical experiences or exhibit (acquired) canonical gaze patterns, such as letters or words.

The fact we found this effect in the right but not the left hemisphere might relate to the well-established right hemisphere dominance for face processing (Le Grand et al., 2003). However, overall decoding performance (across all conditions) was better in left compared with right IOG. Furthermore, the interaction between hemisphere and condition was only significant when tested across the four types of decoding shown in Figure 5, but not when tested for the two main conditions shown in Figure 4. This might be related to a lack of statistical power but shows that the trend for a hemispheric difference of this effect has to be interpreted with caution (Nieuwenhuis et al., 2011). The amplitude data suggest a third, potentially related hemispheric difference: ROIs in the left hemisphere showed a stronger contralateral bias. A possible explanation for this is that right hemisphere ROIs might have had better coverage of the ipsilateral hemifield (Fig. 4B; a result that would echo recent data by Kay et al., 2015, their Figs. 2B and S2A).

The general ability to decode isolated face features differing in retinotopic location is in line with previous studies that have shown that fMRI responses in human FFA and especially OFA are sensitive to stimulus location (Hemond et al., 2007; Kovács et al., 2008; Schwarzlose et al., 2008; Yue et al., 2011; Kay et al., 2015) and that OFA (Pitcher et al., 2007; Liu et al., 2010; Nichols et al., 2010; Henriksson et al., 2015) as well as the N170 (Smith et al., 2004, 2009) seem mainly driven by face parts, not their arrangement. One of our analyses showed that stimulus location could be decoded for a given feature in most ROIs, whereas feature identity at a given location could not be decoded significantly above

chance level. Interestingly, the advantage for typical stimulus locations was due to an improvement over “pure” retinotopic decoding. Previous studies show that cells in the posterior lateral face patch exhibit face feature tuning (Issa and DiCarlo, 2012), that part based face processing evokes BOLD responses and release from BOLD adaptation in FFA (Yovel and Kanwisher, 2004; Lai et al., 2014), and that face components can be distinguished based on BOLD patterns in face-sensitive cortex (Nichols et al., 2010). Against this background, our inability to decode face features at a given location might be due to the short stimulus durations (200 ms) and to the use of a noise mask. The rationale behind this was to stay as close as possible to the recognition experiment, but future studies should probably use stimuli more amenable to decoding. More generally, decoding algorithms can only yield indirect evidence regarding encoding hypotheses and depend on information being represented at a spatial scale that can be picked up across voxels. Recent results show that this is not necessarily the case for some types of face encoding (Dubois et al., 2015). Future studies could use fMRI adaptation (Grill-Spector and Malach, 2001), encoding models (Dumoulin and Wandell, 2008), or a reverse correlation approach (Lee et al., 2013; de Haas et al., 2014) to probe the spatial sensitivity profiles of face-sensitive voxels more directly. They could then directly compare such profiles evoked by different face features at the single voxel level.

Our main hypothesis made no predictions regarding average amplitude levels across an ROI. Depending on integration rule, average amplitude effects could be zero, positive, or negative (see Hypotheses and predictions). Amplitude levels for canonical and reversed feature locations were statistically indistinguishable in left and right IOG as well as left FFA. However, there was an overall trend toward higher amplitudes for reversed stimulus locations that was significant in right FFA. This could be due to, for example, ceiling effects (following a principle of inverse effectiveness). Regardless of the cause for this amplitude effect, it is important to note that it was in the opposite direction of the decoding effect. Thus, increased amplitude levels cannot trivially explain the observed modulation of classification performance.

Together, our results show that canonical gaze behavior and the inherent structure of upright faces result in feature-specific retinotopic biases and recognition performance is tuned to this. We further present neuroimaging data in line with a hypothesized correlation between spatial and feature tuning in right IOG. Our findings have potential implications for mechanistic accounts of the well-established recognition advantage for upright over inverted faces (Yin, 1969) and the effects of face inversion on appearance (Thompson, 1980). In an inverted face, eyes and mouth can never appear at typical locations simultaneously and eye-tracking studies show that they tend to be reversed (Barton et al., 2006; Hills et al., 2012). Therefore, face inversion effects might at least partly be due to violations of feature-specific retinotopic preferences. The summed behavioral effect size of reversed locations amounts to ~60%–80% of that for whole face inversion (Fig. 1E). Isolated feature inversion has effects of similar magnitude. However, the combination of both manipulations seems to increase the effect for mouth stimuli, but surprisingly not for eye stimuli. Future studies should further investigate the role of feature-based effects in face inversion, and particularly that of retinal feature displacement. Finally, our findings might be relevant for conditions affecting face recognition and gaze behavior toward faces such as autism spectrum disorders (Klin et al., 2002; Pelphrey et al., 2002; Weigelt et al., 2012; Jones and Klin, 2013) and congenital prosopagnosia (Behrmann and Avidan, 2005;

Schwarzer et al., 2007). Future studies should investigate whether these differences in gaze behavior are linked to altered spatial distributions of feature sensitivity.

References

- Afraz A, Pashkam MV, Cavanagh P (2010) Spatial heterogeneity in the perception of face and form attributes. *Curr Biol* 20:2112–2116. [CrossRef Medline](#)
- Barton JJ, Radcliffe N, Cherkasova MV, Edelman J, Intriligator JM (2006) Information processing during face recognition: the effects of familiarity, inversion, and morphing on scanning fixations. *Perception* 35:1089–1105. [CrossRef Medline](#)
- Behrmann M, Avidan G (2005) Congenital prosopagnosia: face-blind from birth. *Trends Cogn Sci* 9:180–187. [CrossRef Medline](#)
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436. [CrossRef Medline](#)
- Chan AW, Kravitz DJ, Truong S, Arizpe J, Baker CI (2010) Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nat Neurosci* 13:417–418. [CrossRef Medline](#)
- de Haas B, Schwarzkopf DS, Anderson EJ, Rees G (2014) Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. *Curr Biol* 24:R66–R67. [CrossRef Medline](#)
- Deichmann R, Schwarzbauer C, Turner R (2004) Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 and 3 T. *Neuroimage* 21:757–767. [CrossRef Medline](#)
- Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53:1–15. [CrossRef Medline](#)
- Dubois J, de Berker AO, Tsao DY (2015) Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci* 35:2791–2802. [CrossRef Medline](#)
- Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. *Neuroimage* 39:647–660. [CrossRef Medline](#)
- Edelman S, Intrator N (2000) (Coarse coding of shape fragments) + (retinotopy) approximately = representation of structure. *Spat Vis* 13:255–264. [CrossRef Medline](#)
- Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12:1187–1196. [CrossRef Medline](#)
- Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107:293–321. [CrossRef Medline](#)
- Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34:479–490. [CrossRef Medline](#)
- Hemond CC, Kanwisher NG, Op de Beeck HP (2007) A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS One* 2:e574. [CrossRef Medline](#)
- Henriksson L, Mur M, Kriegeskorte N (2015) Faciotopy: a face-feature map with face-like topology in the human occipital face area. *Cortex* 72:156–167. [CrossRef Medline](#)
- Hills PJ, Sullivan AJ, Pake JM (2012) Aberrant first fixations when looking at inverted faces in various poses: the result of the centre-of-gravity effect? *Br J Psychol* 103:520–538. [CrossRef Medline](#)
- Hsiao JH, Cottrell G (2008) Two fixations suffice in face recognition. *Psychol Sci* 19:998–1006. [CrossRef Medline](#)
- Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002) Image distortion correction in fMRI: a quantitative evaluation. *Neuroimage* 16:217–240. [CrossRef Medline](#)
- Issa EB, DiCarlo JJ (2012) Precedence of the eye region in neural processing of faces. *J Neurosci* 32:16666–16682. [CrossRef Medline](#)
- Jones W, Klin A (2013) Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* 504:427–431. [CrossRef Medline](#)
- Kasiński A, Florek A, Schmidt A (2008) The put face database. *Image Process Commun* 13:59–64.
- Kay KN, Weiner KS, Grill-Spector K (2015) Attention reduces spatial uncertainty in human ventral temporal cortex. *Curr Biol* 25:595–600. [CrossRef Medline](#)
- Kleiner M, Brainard D, Pelli D (2007) What's new in Psychtoolbox-3? *Perception* 36:14.
- Klin A, Jones W, Schultz R, Volkmar F, Cohen D (2002) Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch Gen Psychiatry* 59:809–816. [CrossRef Medline](#)
- Konkle T, Oliva A (2012) A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74:1114–1124. [CrossRef Medline](#)
- Kovács G, Cziráki C, Vidnyánszky Z, Schweinberger SR, Greenlee MW (2008) Position-specific and position-invariant face aftereffects reflect the adaptation of different cortical areas. *Neuroimage* 43:156–164. [CrossRef Medline](#)
- Kravitz DJ, Vinson LD, Baker CI (2008) How position dependent is visual object recognition? *Trends Cogn Sci* 12:114–122. [CrossRef Medline](#)
- Kravitz DJ, Kriegeskorte N, Baker CI (2010) High-level visual object representations are constrained by position. *Cereb Cortex* 20:2916–2925. [CrossRef Medline](#)
- Lai J, Pancaroglu R, Oruc I, Barton JJ, Davies-Thompson J (2014) Neuroanatomic correlates of the feature-saliency hierarchy in face processing: an fMRI-adaptation study. *Neuropsychologia* 53:274–283. [CrossRef Medline](#)
- Lee S, Papanikolaou A, Logothetis NK, Smirnakis SM, Keliris GA (2013) A new method for estimating population receptive field topography in visual cortex. *Neuroimage* 81:144–157. [CrossRef Medline](#)
- Le Grand R, Mondloch CJ, Maurer D, Brent HP (2003) Expert face processing requires visual input to the right hemisphere during infancy. *Nat Neurosci* 6:1108–1112. [CrossRef Medline](#)
- Liu J, Harris A, Kanwisher N (2010) Perception of face parts and face configurations: an fMRI study. *J Cogn Neurosci* 22:203–211. [CrossRef Medline](#)
- Maurer D, Grand RL, Mondloch CJ (2002) The many faces of configural processing. *Trends Cogn Sci* 6:255–260. [CrossRef Medline](#)
- Miner M, Park DC (2004) A lifespan database of adult facial stimuli. *Behav Res Methods Instrum Comput* 36:630–633. [CrossRef Medline](#)
- Nichols DF, Betts LR, Wilson HR (2010) Decoding of faces and face components in face-sensitive human visual cortex. *Front Psychol* 1:28. [CrossRef Medline](#)
- Nieuwenhuis S, Forstmann BU, Wagenmakers EJ (2011) Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat Neurosci* 14:1105–1107. [CrossRef Medline](#)
- Orlov T, Makin TR, Zohary E (2010) Topographic representation of the human body in the occipitotemporal cortex. *Neuron* 68:586–600. [CrossRef Medline](#)
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442. [CrossRef Medline](#)
- Pelphrey KA, Sasson NJ, Reznick JS, Paul G, Goldman BD, Piven J (2002) Visual scanning of faces in autism. *J Autism Dev Disord* 32:249–261. [CrossRef Medline](#)
- Peterson MF, Eckstein MP (2012) Looking just below the eyes is optimal across face recognition tasks. *Proc Natl Acad Sci U S A* 109:E3314–E3323. [CrossRef Medline](#)
- Peterson MF, Eckstein MP (2013) Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychol Sci* 24:1216–1225. [CrossRef Medline](#)
- Pitcher D, Walsh V, Yovel G, Duchaine B (2007) TMS evidence for the involvement of the right occipital face area in early face processing. *Curr Biol* 17:1568–1573. [CrossRef Medline](#)
- Pitcher D, Walsh V, Duchaine B (2011) The role of the occipital face area in the cortical face perception network. *Exp Brain Res* 209:481–493. [CrossRef Medline](#)
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025. [CrossRef Medline](#)
- Schwarzer G, Huber S, Grüter M, Grüter T, Gross C, Hipfel M, Kennerknecht I (2007) Gaze behaviour in hereditary prosopagnosia. *Psychol Res* 71:583–590. [CrossRef Medline](#)
- Schwarzlose RF, Swisher JD, Dang S, Kanwisher N (2008) The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci U S A* 105:4447–4452. [CrossRef Medline](#)
- Silson EH, Chan AW, Reynolds RC, Kravitz DJ, Baker CI (2015) A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *J Neurosci* 35:11921–11935. [CrossRef Medline](#)
- Smith ML, Gosselin F, Schyns PG (2004) Receptive fields for flexible face categorizations. *Psychol Sci* 15:753–761. [CrossRef Medline](#)
- Smith ML, Fries P, Gosselin F, Goebel R, Schyns PG (2009) Inverse mapping

- the neuronal substrates of face categorizations. *Cereb Cortex* 19:2428–2438. [CrossRef Medline](#)
- Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 9:255–266. [CrossRef Medline](#)
- Tanaka JW, Farah MJ (1993) Parts and wholes in face recognition. *Q J Exp Psychol A* 46:225–245. [CrossRef Medline](#)
- Tanaka JW, Sengco JA (1997) Features and their configuration in face recognition. *Mem Cognit* 25:583–592. [CrossRef Medline](#)
- Thompson P (1980) Margaret Thatcher: a new illusion. *Perception* 9:483–484. [CrossRef Medline](#)
- Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci U S A* 105:19514–19519. [CrossRef Medline](#)
- Ullman S (2007) Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn Sci* 11:58–64. [CrossRef Medline](#)
- van Belle G, Ramon M, Lefèvre P, Rossion B (2010) Fixation patterns during recognition of personally familiar and unfamiliar faces. *Front Psychol* 1:20. [CrossRef Medline](#)
- van den Hurk J, Pegado F, Martens F, Op de Beeck HP (2015) The search for the face of the visual homunculus. *Trends Cogn Sci* 19:638–641. [CrossRef Medline](#)
- Vieira TF, Bottino A, Laurentini A, De Simone M (2013) Detecting siblings in image pairs. *Vis Comput* 30:1333–1345.
- Võ ML, Smith TJ, Mital PK, Henderson JM (2012) Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *J Vis* 12:3. [CrossRef Medline](#)
- Weigelt S, Koldewyn K, Kanwisher N (2012) Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neurosci Biobehav Rev* 36:1060–1084. [CrossRef Medline](#)
- Weiner KS, Grill-Spector K (2013) Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychol Res* 77:74–97. [CrossRef Medline](#)
- Werner S, Noppeney U (2011) The contributions of transient and sustained response codes to audiovisual integration. *Cereb Cortex* 21:920–931. [CrossRef Medline](#)
- Yin R (1969) Looking at upside-down faces. *J Exp Psychol* 81:141–145. [CrossRef](#)
- Yovel G, Kanwisher N (2004) Face perception: domain specific, not process specific. *Neuron* 44:889–898. [CrossRef Medline](#)
- Yue X, Cassidy BS, Devaney KJ, Holt DJ, Tootell RB (2011) Lower-level stimulus features strongly influence responses in the fusiform face area. *Cereb Cortex* 21:35–47. [CrossRef Medline](#)