Behavioral/Cognitive

Representations of Facial Identity Information in the Ventral Visual Stream Investigated with Multivoxel Pattern Analyses

Elfi Goesaert and Hans P. Op de Beeck

Laboratory of Biological Psychology, University of Leuven, 3000 Leuven, Belgium

The neural basis of face recognition has been investigated extensively. Using fMRI, several regions have been identified in the human ventral visual stream that seem to be involved in processing and identifying faces, but the nature of the face representations in these regions is not well known. In particular, multivoxel pattern analyses have revealed distributed maps within these regions, but did not reveal the organizing principles of these maps. Here we isolated different types of perceptual and conceptual face properties to determine which properties are mapped in which regions. A set of faces was created with systematic manipulations of featural and configural visual characteristics. In a second part of the study, personal and spatial context information was added to all faces except one. The perceptual properties of faces were represented in face regions and in other regions of interest such as early visual and object-selective cortex. Only representations in early visual cortex were correlated with pixel-based similarities between the stimuli. The representation of nonperceptual properties was less distributed. In particular, the spatial location associated with a face was only represented in the parahippocampal place area. These findings demonstrate a relatively distributed representation of perceptual and conceptual face properties that involves both face-selective/sensitive and non-face-selective cortical regions.

Introduction

The ability to quickly and efficiently recognize a face is one of the most intriguing aspects of human vision and has received a great deal of attention. Despite this interest, the specific mechanisms underlying this ability remain unclear. A number of brain regions have been identified that are said to be involved in face recognition, the best known being the fusiform face area (FFA) and the occipital face area (OFA) (Kanwisher et al., 1997; Kanwisher and Yovel, 2006; Pitcher et al., 2011). Recently, a more anterior part of the temporal cortex has also been implicated in face perception (Kriegeskorte et al., 2007; Nestor et al., 2008; Tsao et al., 2008; Pinsk et al., 2009; Rajimehr et al., 2009; Nestor et al., 2011). Although these regions are typically treated as uniform regions of interest (ROIs), recent studies have used a multivoxel analysis approach to suggest the existence of a distributed pattern of selectivity within these regions. Multivoxel pattern analyses (MVPAs) have been widely applied to reveal distributed selectivity maps that are impossible/difficult to detect with standard fMRI analyses (Haxby et al., 2001; Op de Beeck et al., 2010), and the few studies that have applied this method to face perception so far (Kriegeskorte et al., 2007; Nestor et al., 2011) have found evidence for distinctive response patterns within some of these face regions.

However, the findings do not match up completely and, most importantly, the studies did not reveal the face characteristics that underlie these distributed selectivity maps. First, the regions showing a distributed map differed between the studies and sometimes these regions are not purely face selective (Nestor et al., 2011). Second, although one study (van den Hurk et al., 2011) suggested that the map in FFA, typically considered to be a visual region, is also sensitive to nonvisual, associative properties of faces, another study (Nestor et al., 2011) suggested that even the maps more anterior in the temporal lobe still mostly reflect visual properties. These conclusions are mostly speculative given that neither study explicitly isolated different face properties.

Here we isolated specific face characteristics. First, we compared different perceptual dimensions on which faces can vary: featural changes (e.g., the shape of the mouth, color of the eyes, etc.) versus configural changes (the spacing between the face parts). Second, we investigated how and where contextual information is processed. Background information was added to most of the faces in a second part of the study by providing a name, occupation, and location. Our results indicate that the distributed maps in face regions of the ventral visual stream (OFA, FFA, and an anterior face-sensitive patch, aIT), as well as in other visual regions, mostly reflect visual differences among faces. In addition, we found that the map in the face regions and early visual areas are also sensitive to whether a face is associated (by training) with nonvisual context and that only the parahippocampal place area is sensitive to which spatial context is associated with which face.

Materials and Methods

Subjects. Twelve subjects (four female), ages 22–33 years, participated in the fMRI study. They were all right handed and had normal or corrected-to-normal vision. Written informed consent was acquired from all sub-

jects before each scan session, and the study was approved by the Committee for Medical Ethics of the University of Leuven (Leuven, Belgium).

Stimulus construction. Nine faces were constructed, one "base face" and eight exemplar faces that differed equally from the base face. The faces were constructed using Facegen Modeler software (Singular Inversions). Each of these eight faces consisted of one of two possible feature sets. Feature set 1 had blue eyes, light eyebrows, and thick lips; feature set 2 had brown eyes, dark eyebrows, and thin lips. Subsequently, in each of the eight faces, eye position and mouth position were manipulated to induce configural differences. Eyes were either low and close together or high and further apart and the mouth was either raised or lowered compared with the base face (Fig. 1a). The behavioral sensitivity for featural and configural differences was tested in a series of small behavioral pilot experiments to ensure a similar sensitivity, converging on the final set of faces with comparable similarity of the differences induced by configural versus featural changes.

Behavioral training sessions. During training, all faces except the base face were associated with personal and spatial context information. The eight trained faces were divided into two groups that were associated with one of two locations, either a bar or train station. These two groups were allocated according to a rule that was orthogonal to the featural and configural dimensions: each group consisted of intermingled feature sets and configurations to prevent subjects from relating the background information to one of these types of visual information. Figure 1b provides a schematic representation of this division. The training sessions were implemented on a PC with a CRT monitor (resolution 1024 × 768, refresh rate 75 Hz).

There were two parts to the behavioral training: a first session only provided the information associated with each face, whereas in the second session, a training procedure was constructed to fully acquaint subjects with the background information linked with the faces. In the first session, a location (scene image) was presented (1 s), followed by the presentation of a face next to this location. Underneath the stimuli, background information of the presented face was shown: a name (Sam, Pieter, Bob, Walter, Mark, Frank, Simon, or Jan), together with the activity the person was performing at that location (at the bar: takes your order, delivers your drink, asks for a light, spills his drink on you; at the station: sells you your train ticket, asks for the time, checks your train ticket, asks to borrow a pen). Subjects pushed a button as soon as they felt they had processed the information sufficiently and were then shown a quick overview before moving on to the next trial (face + all information: name-location-activity, lasting 2 s). During the second training session, a trial started with a face shown for 1.5 s, together with a question about the person's name, activity, or location and two possible response alternatives. After the subject answered, feedback was given (correct or incorrect), including another presentation of the face and a summary of the correct information (1 s). Training took place at most 2 d before the final scan session and subjects were considered fully trained if they performed the training task with an accuracy of 95% or higher in 3 consecutive blocks of trials (48 trials per block). Subjects needed on average ~six blocks of trials to complete the training. The day of the scan session, subjects again performed the training task to ensure that performance was still high (95% or higher). During this last training session, subjects performed on average two more blocks of the training.

General fMRI scanning procedure. Subjects participated in three scan sessions: two before training (performed on two different days) and one after training (performed on a third day). The first two sessions each encompassed nine experimental runs with the face stimuli and three to four localizer runs. The last session consisted of 15 experimental runs.

All experiments were programmed using MATLAB (Mathworks) and PsychToolbox (Brainard, 1997) for stimulus presentation and response
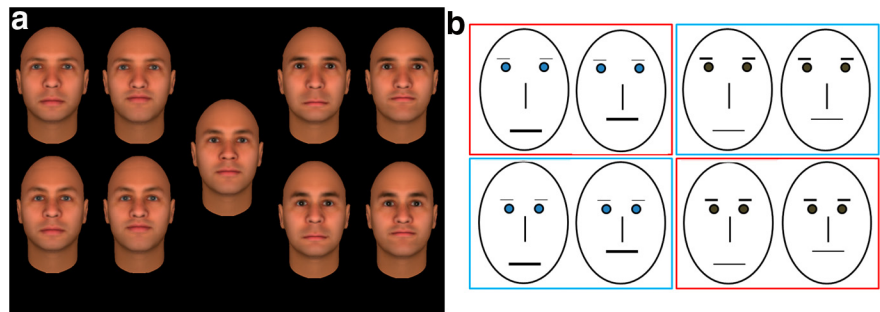
Figure 1. Stimuli and overview of context categories. *a*, Stimuli used in the fMRI study. The base face is presented in the middle, the left side of the panel consists of faces belonging to feature set 1 (blue eyes, light eyebrows, thick lips), and the right side shows the faces of feature set 2 (brown eyes, dark eyebrows, thin lips). For each feature set, eyes could be up and apart (top) or down and close together (bottom) and, for each of these eye positions, the mouth was either down (left) or up (right). *b*, Schematic representation of the stimuli and division into context locations. Differences between the eight trained faces are schematically represented and are organized similar to *a*. The faces were divided into location groups (red and blue squares for bar and station, respectively) orthogonally to the featural and configural differences: each group contained the same amount of members of each feature set and of each value of the configural dimension.

registration. During the fMRI experiment, the stimuli were presented on a 30 inch radiofrequency-shielded LCD screen (resolution 1044 × 900, refresh rate 60 Hz) at the back of the scanner, which subjects viewed via a mirror on top of the head coil. The viewing distance to the screen was ~125 cm (eyes to mirror: 11 cm; mirror to screen: 114 cm).

fMRI experimental runs. There were three blocks of each of the nine faces. These blocks had a variable length lasting between 4 and 8 s per block. Stimuli were shown for 0.5 s, with a 0.5 s interstimulus interval, and their position on the screen was varied (the maximum offset from fixation point was ~1 degree). The size of the stimuli on the screen was ~4 (horizontal) by 6 degrees (vertical). An experimental run also contained four fixation blocks that lasted 12 s each intermingled with the face blocks. The order of the conditions was pseudorandomized within runs and the order of the runs was counterbalanced across participants. The block length was pseudorandomized within and counterbalanced between runs to ensure that across the entire scan session, each stimulus was shown an equal amount of time and each scan sequence was equally long (54 s for one part of the sequence containing all possible face blocks).

The task was different in the pretraining scan sessions compared with the posttraining sessions. Before training, subjects were asked to press a button each time a different face appeared (i.e., at the start of each block of a different condition). Before the first session, subjects practiced the task outside of the scanner to ensure that the differences between the different faces were clear to subjects. Subjects had to perform this task on average two times to get at a performance rating of ~80% or more, meaning that the differences between the nine faces were clear to them without the need for any further training. After training, subjects were asked to again press a button whenever the presented face changed, but had to press one of two buttons depending on the location the face belonged to (bar or station). In the case of the base face, which was not trained, subjects were asked to push a button corresponding to the location the face could belong to or push a button at random if they had no preference. As an addition to the training sessions (see previous), subjects were asked to perform this task in advance to ensure that they could perform the task correctly (on average, two sequences of 27 experimental blocks with an average performance of 92%).

fMRI localizer runs. In addition to the experimental runs, a number of independent localizer runs were collected. During the first session, subjects were presented with blocks of grayscale faces, objects, and scrambled objects and performed an odd-man-out task on size. Each block of stimuli contained 20 different images of the category. There were three fixation blocks lasting 15 s at the start, middle, and end of each run. In between, 12 blocks of the stimulus categories (four blocks per category) were presented lasting 15 s each, and the presentation order of all blocks was counterbalanced across runs. The first three subjects completed three runs of this localizer; the other subjects four runs. During the

second session, all subjects performed four runs of the same localizer with an additional category, scenes (yielding 4 × 4 stimulus blocks and 5 fixation blocks per run), and subjects were asked to perform a one-back task (they pressed a button whenever the stimulus was identical to the one shown in the preceding trial).

fMRI scanning parameters and preprocessing. fMRI data were collected using a 3T Philips Intera magnet (Department of Radiology, University of Leuven) with a 32-channel SENSE head coil with an echo-planar imaging sequence (70 time points per time series or "run" for the main experiment, 75 time points for localizer 1, and 105 time points for localizer 2; repetition time, 3000 ms; echo time, 29.8 ms; acquisition matrix 104 × 104, resulting in a 2.0 by 2.0 mm² in-plane voxel size; 47 slices oriented approximately halfway between a coronal and horizontal plane and included most of the brain except the most superior parts of frontal and parietal cortex, with slice thickness 2 mm and interslice gap 0.2 mm). For most subjects, a T1-weighted anatomical image was already available; for five subjects this was acquired during the first or second scan session (resolution 0.98 × 0.98 × 1.2 mm, 9.6 ms TR, 4.6 ms TE, 256 × 256 acquisition matrix, 182 coronal slices).

All data were preprocessed using SPM8 (Department of Cognitive Neurology, Wellcome Trust, London). During preprocessing, the data were corrected for differences in acquisition time, motion corrected, and the anatomical image was realigned to the functional data. The data were normalized to MNI space using the coregistered segmented anatomical image and voxels were resampled to a voxel size of 2 × 2 × 2 mm. In subjects for whom problems arose in the segmentation process, the data were normalized to MNI space using a T1 template image. Finally, functional images were smoothed with a 4 mm full-width at half maximum Gaussian kernel. ROIs were defined using the SPM8 contrast manager and a custom-made MATLAB script for the selection of significantly activated voxels displayed on coronal sections.

Localization of ROIs. Using all localizer runs, the face-selective ventral regions were defined by a contrast of faces versus objects, with a standard threshold of $p < 0.0001$ uncorrected that was more lenient if <20 voxels could be identified for FFA (at 0.001 uncorrected for one subject) or OFA (once at 0.05 uncorrected, once at 0.01 uncorrected, and once at 0.001 uncorrected). In the fusiform region, we only selected the most anterior region as FFA (in the middle fusiform region); likewise, only the most posterior part of OFA was selected. This was aimed at including mainly the middle fusiform gyrus area of FFA (Weiner and Grill-Spector, 2010) and avoid possible overlap between different face-selective regions. aIT consisted of face-sensitive regions anterior to FFA and was defined by a faces versus baseline contrast at $p < 0.001$ uncorrected to ensure that a large enough region was selected (Mur et al., 2010). The average MNI coordinates for the aIT ROI were ± 33 −8 −33.

These aIT activations span both the AFP1 and AFP2 regions as labeled by Tsao et al. (2008). Our data do not allow differentiation between AFP1 and AFP2 because seven subjects showed only one region and, in several cases, we were not sure about the determination of a patch as being either AFP1 or AFP2. In Tsao et al. (2008), the exact anatomical position of these regions varied among individual subjects; furthermore, Rajimehr et al. (2009) only distinguished one large anterior temporal face patch). Therefore, we collapsed the data of all these anterior regions into one ROI, making the untestable assumption that both patches contained similar representations.

The average size of the face regions was as follows: FFA, 249 voxels (SD = 178.88); OFA, 101 voxels (SD = 101.12); and aIT, 151 voxels (SD = 79.28).

Parahippocampal place area (PPA) was defined by a scene versus objects contrast, thresholded at $p < 0.0001$ uncorrected (M = 715 voxels, SD = 536.45). The lateral occipital complex (LOC) consisted of lateral occipital and posterior fusiform sulcus, defined by an objects versus scrambled contrast at $p < 0.0001$ uncorrected and excluding all face- or scene-selective voxels defined by the faces versus objects and scenes versus objects contrasts at $p < 0.05$ uncorrected (M = 247 voxels, SD = 117.98). An early visual cortex area (EVC) was defined by an all (objects, faces, scrambled, scenes) versus baseline contrast at $p < 0.0001$ uncorrected and using an anatomical mask that selected Brodmann area 17, which approximately corresponds to V1 constructed with PickAtlas
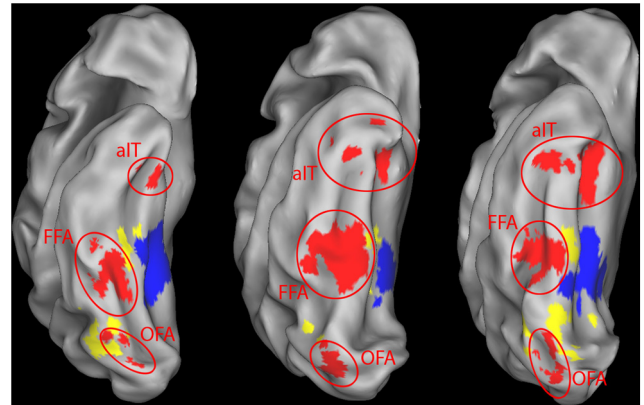

Figure 2. Functionally defined ROIs shown on an inflated view of the right hemisphere of three subjects. Three types of regions investigated in this study are shown, defined as described in the Materials and Methods section: face regions (red), scene-selective regions (PPA, blue), and object-selective regions (LOC, yellow). The face regions consist of OFA, FFA, and aIT.

(M = 246 voxels, SD = 55.12; Advanced NeuroScience Imaging Research Laboratory, Wake Forest University, Winston-Salem, NC). For further control analyses on potential motor confound associated with the task performed in the scanner, we also determined primary motor cortex (M1) based upon an all versus baseline contrast at $p < 0.01$ uncorrected and an anatomical mask that consisted of Brodmann area 4, again constructed with PickAtlas. M1 was completely outside the scanned volume for four subjects, so it was only determined for eight subjects. Finally, we defined a white matter control region consisting of voxels that were mainly situated around the corpus callosum, away from any possibly relevant gray matter regions and not responsive to an all versus baseline contrast (M = 226 voxels, SD = 96.22). Figure 2 shows an inflated ventral view of the right hemisphere of three participants constructed from their anatomical image using CARET software (Van Essen et al., 2001). The face-, scene-, and object-selective areas are marked in color (OFA, FFA, aIT, PPA, and LOC). Although only one hemisphere is shown in the figure, all ROIs selected with the localizer data were bilateral regions.

MVPA of face identity. For MVPA, the experimental sessions were divided into a before-training and after-training dataset. The preprocessing of data from different datasets was performed separately. Alignment of the data was secured (and visually inspected afterward) through the normalization step, which allowed us to, for example, use ROIs defined using data from one dataset to analyze data from another dataset. A GLM was estimated separately for each of the two datasets using the preprocessed data. The model consisted of nine conditions of interest, one for each of the faces in the study and six covariates (the translation and rotation parameters that were calculated during realignment).

After fitting the GLM, the parameter estimates (β values) for all voxels in the model were used as input for the MVPA analysis, yielding a pattern of activation for each condition in each run. These parameter estimates were standardized to a mean of zero and a variance of one. The data were further analyzed with linear support vector machines (SVMs) using the OSU SVM MATLAB toolbox (*www.sourceforge.net/projects/svm/*) and the same methodology as in previous studies (Op de Beeck et al., 2008; Op de Beeck et al., 2010). The SVMs were trained and tested on pairwise discrimination of two face conditions. The data, in total consisting of nine response patterns (one per condition) per run, were divided randomly into two parts, one used to train the SVM classifier and one to test it. The pretraining dataset was divided into 15 training and three test runs, and the posttraining dataset was divided into 12 training and three test runs. The SVM classifier was trained and tested in pairwise classification across 100 random repetitions with different divisions into training and test runs. Last, as described previously by Nestor et al. (2011), the data were converted to a sensitivity index (the d′ index from signal detection theory). For this conversion, the decoding performance (after averaging across all relevant pairwise comparisons) was split into hits and false alarm proportions, which were z-transformed. Finally, z(false

alarms) was subtracted from z(hits). We checked for differences in mean response strength between the conditions compared with MVPA, and no differences were noted unless noted otherwise in the Results section.

Given our cross-validation procedures (e.g., using data from different runs for training and test), the sensitivity that is expected by chance with random data is zero. As an additional control, we also applied our MVPA procedure using the response patterns of the white matter control region. We performed all the statistical tests done on an individual region on this white matter ROI, including ANOVA. The total number of these tests was 22 and the average sensitivity of these tests was 0.015 (SEM = 0.085), thus providing an extra-empirical confirmation of our theoretical expectation of a chance sensitivity of zero.

MVPA of the spatial location associated with scenes. All methods were the same as for the analysis of face identity except as follows. For the analyses of the coding of the spatial location associated with faces, we applied a GLM with only three conditions of interest: (1) trained faces associated with the bar, (2) trained faces associated with the station, and (3) the untrained base face. SVM classifiers were trained and tested on the distinction between the first and second conditions.

We also performed an extra control analysis to determine whether the coding of spatial location in PPA (see Results) could reflect the motor response of subjects. Indeed, response-button assignment was not reversed during the posttraining scan session, and thus the associated scene was confounded with the button to press. For a control, we took advantage of the fact that subjects were allowed to choose which button to press for the untrained base face. For this base face, we then had a button press without any association with a scene. Across subjects, we had many base face blocks in which subjects pressed the same button as for the scene-associated faces and many base face blocks in which the button for the bar-associated faces was pressed.

The control analysis for the response confound was based upon a GLM with four conditions of interest: (1) trained faces associated with the bar, (2) trained faces associated with the station, (3) the untrained base face with the button pressed that corresponds to the bar, and (4) the untrained base face with the button pressed that corresponds to the station. We then trained SVM classifiers on the distinction between conditions (1) and (2) and tested them on the distinction between conditions (3) and (4). Because conditions (3) and (4) correspond to conditions (1) and (2) in terms of motor response, but not in terms of the scene associated with faces during training, a successful generalization to conditions (3) and (4) would be evidence that a region would represent the motor response rather than the associated scene.

Results

Representations of faces in face-selective and face-sensitive areas

General decoding of face identity and perceptual information

The first step in the investigation of face identity information representations in the ventral visual stream was to see whether the face regions actually represent any face information. To test this hypothesis, we averaged the SVM decoding sensitivity of all face pairs across the pretraining and the posttraining datasets and across the three face regions (OFA, FFA and aIT), considering these regions together as one "joined face region." Averaged across all face pairs, we found significant decoding (2-tailed paired t test for the difference from a sensitivity of zero, averaged across datasets: sensitivity = 0.087, SEM = 0.025, $t_{(11)}$ = 3.48, p = 0.0052).

In further analyses, we focused specifically upon shape pairs in which the faces differed only in one aspect, either configural or featural differences, meaning both types of perceptual information were grouped together as one category. Averaged across all those shape pairs, we again found a significant decoding in the joined face area (2-tailed paired t test for the difference from a sensitivity of zero, averaged across the pretraining and the post-training datasets: $t_{(11)}$ = 4.11, p = 0.0017; Fig. 3). When taking the individual face areas apart, this significant decoding was
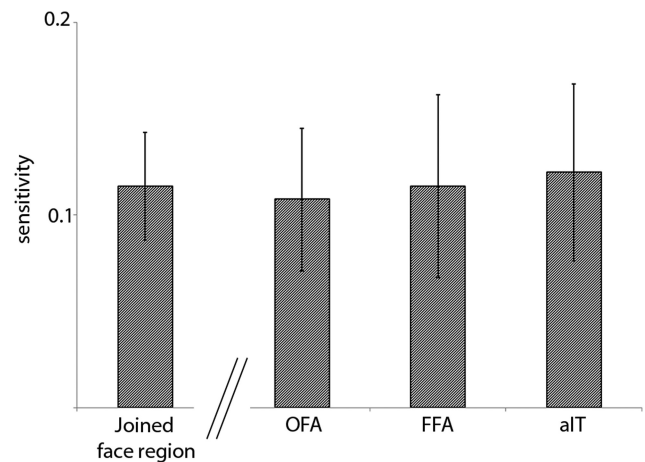

Figure 3. Sensitivity of SVM classifiers to perceptual differences (averaged across featural and configural differences) for the joined face region and the three face regions separately (OFA, FFA, aIT). Error bars display the SEM across subjects.

found in each of the three face regions (2-tailed paired t test for the difference from a sensitivity of zero, averaged across sessions: $t_{(11)}$ = 2.42 p = 0.034 for FFA, $t_{(11)}$ = 2.92 p = 0.014 for OFA, and $t_{(11)}$ = 2.67 p = 0.022 for aIT; Figure 3). When the data were FDR corrected for the three regions, these effects remained significant.

Taking the configural and featural differences apart, the joined face region showed significant decoding for both configural (2-tailed paired t test for the difference from a sensitivity of zero, averaged across sessions: $t_{(11)}$ = 4.56, p = 0.00081) and featural information ($t_{(11)}$ = 3.20, p = 0.0084; Fig. 4a). No difference was found between configural and featural information ($t < 1$, ns). The decoding ability for the three regions for configural and featural information is shown in Figure 4b. To investigate the similarities and differences of the three face regions for the two perceptual dimensions, we performed a 3 × 2 repeated-measures ANOVA. There was no difference in decoding ability between the three face regions ($F < 1$, ns), no difference between configural or featural information ($F < 1$, ns), and no interaction between these two dimensions ($F < 1$, ns). Note that further specific *post hoc* testing of each perceptual dimension (configural and featural) in each of the three face regions did not often yield a significant decoding after correction for multiple comparisons. Therefore, although the aforementioned significant tests allow us to say that the joint activity pattern in the three face regions contains information about each of the dimensions, and that each of the three regions contains information about perceptual characteristics of the faces, our methods are not sensitive enough to uncover reliable information about each perceptual dimension in each face region. For that reason, we should also be cautious about our inability to find differences between the 3 regions in the 3 × 2 ANOVA, because this null result could also be related to the low sensitivity of our data when they are taken apart at that level of detail.

We averaged across the pretraining and posttraining datasets for all of these analyses. Further analyses of the configural and featural dimensions with the inclusion of pretraining versus posttraining as an additional factor did not show any further effects. Specifically, a 2 × 2 repeated-measures ANOVA with pretraining versus posttraining and configural versus featural as factors did not reveal a main effect of pretraining versus posttraining ($F_{(1,11)}$ = 1.28, p = 0.28) and no interaction ($F < 1$, ns) in the joined face region.
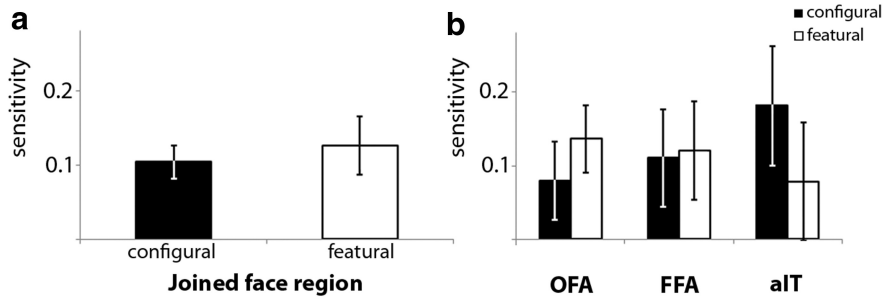
Figure 4. Sensitivity of SVM classifiers to configural and featural differences. *a*, Sensitivity of SVM classifiers to configural and featural differences to the joined face region. *b*, Sensitivity of SVM classifiers to featural and configural differences for each face region separately (OFA, FFA, aIT). Error bars display the SEM across subjects.
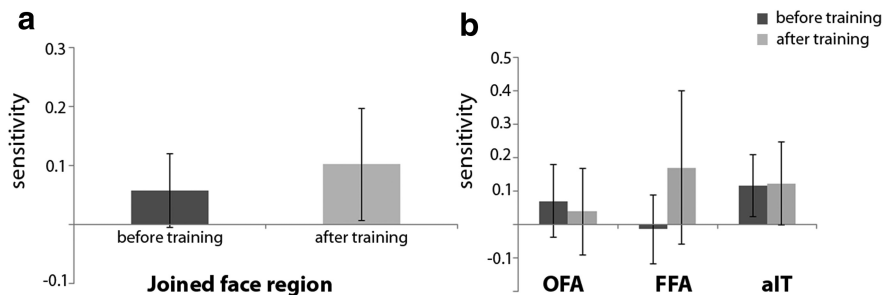

Figure 5. Sensitivity of SVM classifiers to the spatial location associated with faces during training. Sensitivity to spatial location is shown before and after training in the joined face region (*a*) and the three face regions separately (*b*). Error bars display the SEM across subjects.

Context information: which context/location is associated with which face?

During the training session, the eight exemplar faces were divided into two groups corresponding to two possible locations: four faces associated with the bar, and four faces associated with the station (this gives a three-condition GLM of bar faces, station faces, and the base face; see Materials and Methods). Note that, on average, these two conditions were equated in terms of face features and face configuration (i.e., the same number of faces in each group with each feature set so there were no systematic differences between the two conditions in terms of visual dimensions). Therefore, before training, we did not expect to find a significant decoding of the difference between the two conditions. After training, it should be possible to decode the difference between the two conditions, at least if the spatial selectivity pattern in a region conveys information about which context/location is associated with a face. Figure 5a shows the sensitivity of the classifier before and after training in the joined face region. No significant decoding was found after training (2-tailed paired *t* test for the difference from a sensitivity of zero: $t_{(11)} = 1.069, p = 0.31$), nor was there a significant difference between before and after training ($t < 1$, ns). No significant effects were noted when assessing the three face regions separately (Fig. 5b). In general, the spatial selectivity pattern in the face areas of the ventral visual stream does not reflect the spatial context associated with faces.

Familiarity effects concerning the base face

The base face differs from the other eight faces in several aspects. Before and after training, the base face is the most prototypical face from a perceptual point of view—it is the "norm" of this face space (Loffler et al., 2005; Leopold et al., 2006). After training (but not before), the base face differed from the other eight faces in two additional aspects. First, the base face was not trained so

subjects were not exposed to the face for the same amount of time. Second, the base face was not associated with any contextual information. Figure 6a shows the results for comparisons between face pairs in which the base face was compared with the other eight faces before and after training and, again, the results were first investigated for the face regions joined together. Before training, at a time in which only perceptual differences existed between the base face and the other faces, the base face could not be distinguished successfully in the joined face region (2-tailed paired *t* test for the difference from a sensitivity of zero: $t_{(11)} = 1.61, p = 0.14$).

After training, a general increase in decoding ability was apparent, and the joined face region showed significant decoding of the difference between the untrained base face versus the trained faces (2-tailed paired *t* test for the difference from a sensitivity of zero: $t_{(11)} = 4.036, p = 0.0020$), as well as a significant difference between sessions ($t_{(11)} = -3.44, p = 0.0055$). We performed further analyses with face ROI (OFA, FFA, aIT) as a third factor in addition to pretraining versus posttraining and base face versus trained faces (decoding ability for base face vs other faces before and after training is shown in Fig. 6b). There were no differences between ROIs ($F_{(1,11)} = 2.27, p = 0.13$), nor did this factor interact with the other factors ($F_{(1,11)} = 1.53, p = 0.24$). In sum, training was associated with an increased differentiation in the response pattern associated with trained faces versus the untrained base face in ventral face regions.

Given the many aspects on which the base face and the trained faces differ, it is not unlikely that some of these differences might already result in a difference in the mean activation of the ROIs. Therefore, the mean activation values (average β across all voxels in the ROI) were calculated for each face region before and after training for the untrained base face and the other, trained faces and averaged again to get a measure of response strength in the joined face region. We did not find any differences between the base face and the other faces in terms of response strength in the joined face region after training ($t_{(11)} = 1.71, p = 0.12$), which one might have expected based upon the data of Loffler et al. (2005), but also see Davidenko et al. (2012). A 2 × 2 repeated-measures ANOVA was performed for the mean activation in the joined face region before and after training and for the base face and the trained faces. There was a stronger response for the base face across sessions (response strength base face = 1.00, SEM = 0.12, response strength other faces = 0.96, SEM = 0.12, $F_{(1,11)} = 4.99, p = 0.047$). No effect of training was found ($F < 1$) and no interaction was found between training and type of face ($F_{(1,11)} = 1.07, p = 0.32$). Due to the differences in decoding ability in terms of sensitivity before and after training, this small effect in which response strength for the base face was higher than for the other faces is unlikely to have strongly influenced the strong decoding after training.

Scene-selective cortex as a possible source for context information

PPA was included as an a priori ROI due to its possible function as a region involved in processing context information in general (Bar and Aminoff, 2003, but see Epstein and Ward, 2010) and spatial information in particular (Kravitz et al., 2011). Given that our training protocol included associating faces with context information in general and spatial information in particular, activity patterns with PPA might have been modulated by our training.

Before investigating the specific links with context and location information, we looked at the ability of the region to decode face information. Averaging across all face pairs and across the pretraining and the posttraining datasets, we found that the PPA was capable of decoding face identity information (sensitivity = 0.15, SEM = 0.03, 2-tailed paired t test for the difference from a sensitivity of zero, averaged across datasets: $t_{(11)} = 4.88$, $p = 0.00049$). Looking specifically at the perceptual dimensions of configural and featural information, sensitivity to these face pairs was significant for configural information (2-tailed paired t test for the difference from a sensitivity of zero, averaged across sessions: $t_{(11)} = 5.28$, $p = 0.00026$) and marginally significant for featural information (2-tailed paired t test for the difference from a sensitivity of zero, averaged across sessions: $t_{(11)} = 2.038$, $p = 0.066$). In contrast to the face regions, a significant difference was found between configural and featural information ($t_{(11)} = 2.70$, $p = 0.021$; Fig. 7). The effects of configural and featural information before and after training were assessed using a 2 × 2 repeated-measures ANOVA. This yielded no significant training effects ($F < 1$) and no interaction between training and type of perceptual information ($F < 1$).

Next, we investigated whether the pattern of activity in PPA conveyed information about the scene with which the faces were associated using the data from the three-condition GLM analysis described before (Fig. 8a). Before training, there was no sensitivity for the scene with which faces were associated (there was even a tendency toward a negative value, 2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = -2.038$, $p = 0.066$). After training, PPA showed a successful decoding of the two different locations (2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = 2.79$, $p = 0.018$), as well as a significant difference between these conditions before versus after training ($t_{(11)} = -3.97$, $p = 0.002$). In contrast to the absence of any decoding of the associated scene in face regions, these results suggest that spatial context information is present in PPA.

These previous findings are supported by a different type of analysis performed based on the general nine-conditions GLM in which each face was considered a condition (see also Materials and Methods). In this model, face pairs can be grouped according to the location they belong to (same location or different location; Fig. 8b). When considering the effects of the added location information, an interaction between training (before or after training) and location information (face pairs of different vs same locations) was found when performing a 2 × 2 repeated-measures ANOVA ($F_{(1,11)} = 12.68$, $p = 0.0045$). No effects where found for training ($F < 1$) or location ($F < 1$). Investigating the results of PPA in more detail, sensitivity to the conditions after training was significant only for the decoding of different loca-
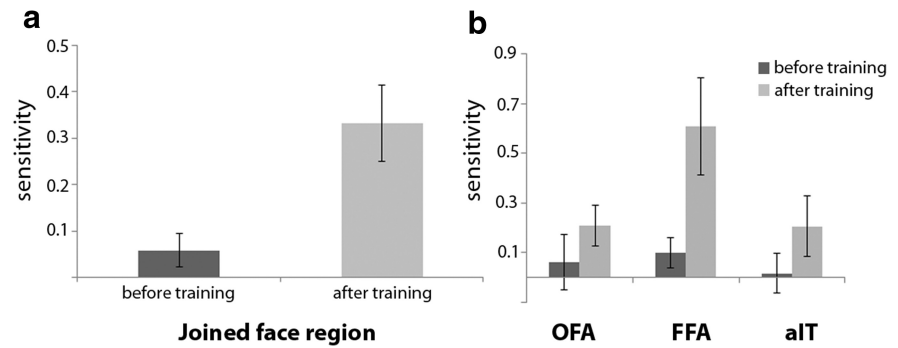

Figure 6. Sensitivity of SVM classifiers to the difference between a trained face and the untrained base face. *a*, Sensitivity of the SVM classifier to the difference between a trained face and the untrained base face before and after training in the joined face region. *b*, Sensitivity of the SVM classifier to the difference between a trained face and the untrained base face before and after training in the face regions separately (OFA, FFA, aIT). Error bars display the SEM across subjects.
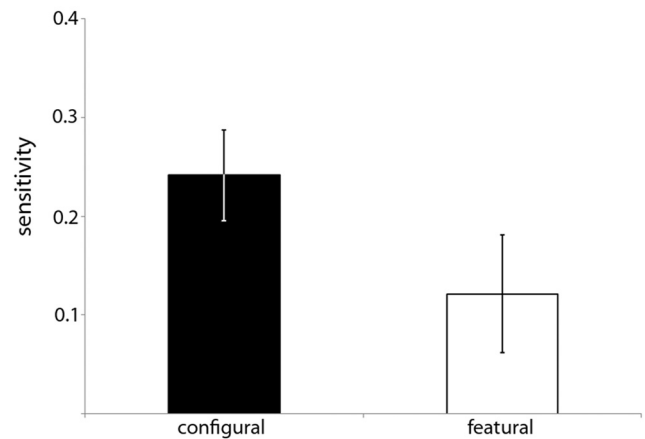

Figure 7. Sensitivity of SVM classifiers to configural and featural differences in PPA. Error bars display the SEM across subjects.

tions (2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = 3.20$, $p = 0.0085$) and not for same locations ($t < 1$, ns). Last, the difference in sensitivity between face pair groups of different and same locations after training was also significant ($t_{(11)} = 2.64$, $p = 0.023$).

The associated scene was confounded with the button that subjects had to press given the task performed during the post-training scans. Could these button presses and the related motor activity, without any association with a scene, be responsible for the apparent coding of scene associations? To answer this question, we divided the base face blocks according to the response made by subjects (for the base face, subjects had the choice to press whatever button they wanted), resulting in a four-condition GLM analysis (see Materials and Methods). We investigated whether classifiers trained to distinguish the bar-associated faces from the station-associated faces could distinguish the bar-button-pressed base face from the station-button-pressed base face. There was no sensitivity for the button pressed in PPA (2-tailed paired t test for the difference from a sensitivity of zero: sensitivity = -0.33, SEM = 0.22, $t_{(11)} = -1.5$, $p = 0.16$). Furthermore, in those subjects where at least part of M1 was included in the scanned volume ($n = 8$), the sensitivity to the button pressed in PPA was significantly smaller than the sensitivity in M1 ($t_{(7)} = 2.79$, $p = 0.027$). In sum, there was a significant sensitivity in PPA for the associated scene without any coding of the response button associated with this scene during scanning.
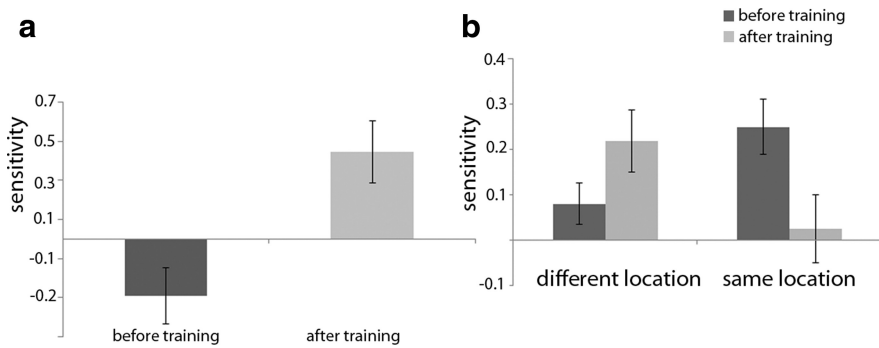
a

b

Figure 8. Sensitivity of SVM classifiers in PPA to the spatial location associated with faces during training. **a**, Sensitivity of SVM classifiers to the difference in multivoxel pattern between the faces associated with the bar location and the faces associated with the station location. Error bars display the SEM across subjects. **b**, Sensitivity of SVM classifiers to the ability in distinguishing faces that differ or do not differ in the associated spatial location. Error bars display the SEM across subjects.
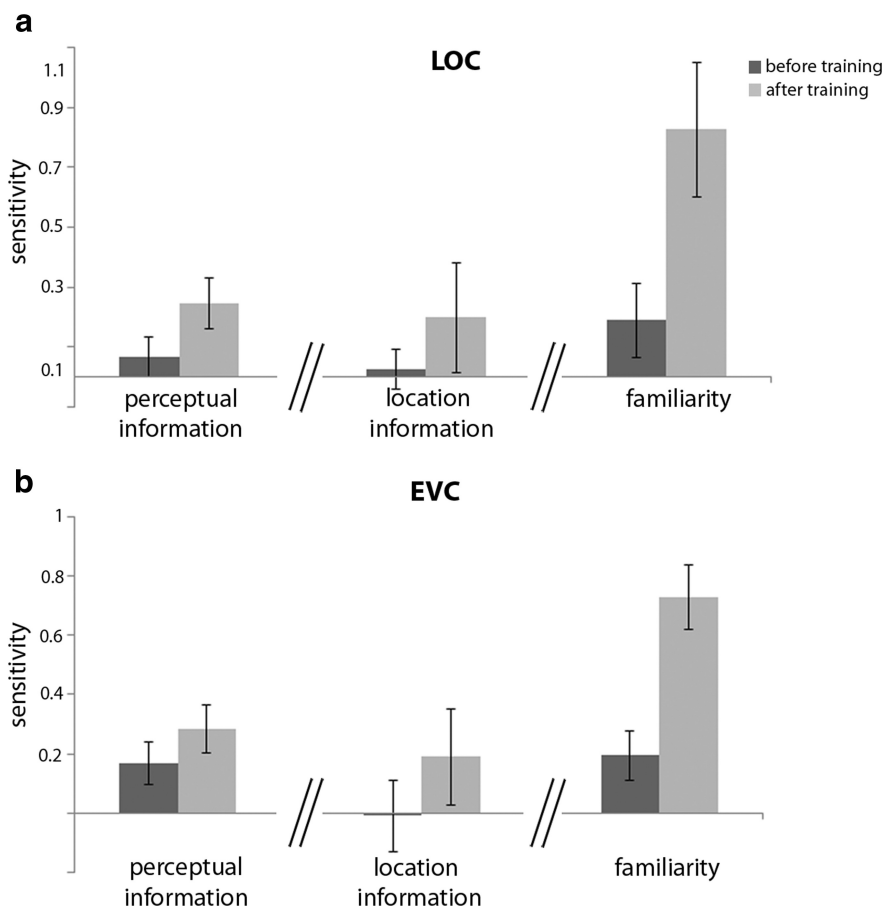
a

b

Figure 9. Sensitivity of SVM classifiers using the patterns of selectivity in LOC (**a**) and EVC (**b**). Sensitivity is shown to the relevant comparison dimensions in this study (perceptual information containing both configural and featural differences, contextual comparisons containing location information based on faces from different locations, and differences between trained and untrained faces). Error bars display the SEM across subjects.

Context information may also play a role in the comparison of the untrained base face with the trained faces. When looking at the SVM performance for this specific comparison before and after training, we found no significant decoding for the base face versus the other, trained faces (sensitivity = 0.067, SEM = 0.078, 2-tailed paired t test for the difference from a sensitivity of zero: $t < 1$, ns) before training. After training, only a marginally significant

effect was found (sensitivity = 0.27, SEM = 0.14, $t_{(11)} = 1.94$, $p = 0.079$), and no difference between the base face comparisons before and after training was found ($t_{(11)} = -1.50, p = 0.16$). This suggests that the context effects in PPA are mostly limited to the context information that contains spatial location information.

Face representations in other regions
In a final step, we analyzed two control regions, LOC, a non-face-selective, shape-selective region in which we did not a priori expect any sensitivity for differences among faces, and EVC. In LOC, there was a significant decoding of face identity across all shape pairs and across pretraining and posttraining datasets (sensitivity = 0.19, SEM = 0.046, 2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = 4.10, p = 0.0018$). We implemented a 2 × 2 repeated-measures ANOVA including the factors pretraining versus posttraining and perceptual dimension (featural versus configural). A marginally significant effect for training was found ($F_{(1,11)} = 3.38, p = 0.093$); no difference between configural and featural information ($F < 1$, ns) and no interaction between training and type of perceptual information ($F_{(1,11)} = 1.03, p = 0.33$) were found. Due to this lack of difference between the two categories, average decoding ability for configural and featural information was taken together and considered as one category representing perceptual information of faces. There was no significant decoding for perceptual information before training (2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = 1.028, p = 0.33$). After training, there was significant decoding ($t_{(11)} = 2.95, p = 0.013$; Fig. 9a). In LOC, there was no representation of the scene associated with the faces before training (2-tailed paired t test for the difference from a sensitivity of zero: $t < 1$, ns) or after training ($t_{(11)} = 1.067, p = 0.31$; Fig. 9a).

In LOC, the classifier was successfully able to distinguish between the base face and the eight other faces (2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = 3.51, p = 0.0049$) after but not before training ($t_{(11)} = 1.54, p = 0.15$). However, the difference in decoding ability before and after training failed to reach significance ($t_{(11)} = -2.048, p = 0.065$; Fig. 9a).

Finally, in EVC, there was a significant decoding of face identity across all shape pairs and across pretraining and posttraining datasets (sensitivity = 0.30, SEM = 0.055, 2-tailed paired t test for the difference from a sensitivity of zero, averaged across datasets: $t_{(11)} = 5.49, p = 0.00019$). A 2 × 2 repeated-measures ANOVA
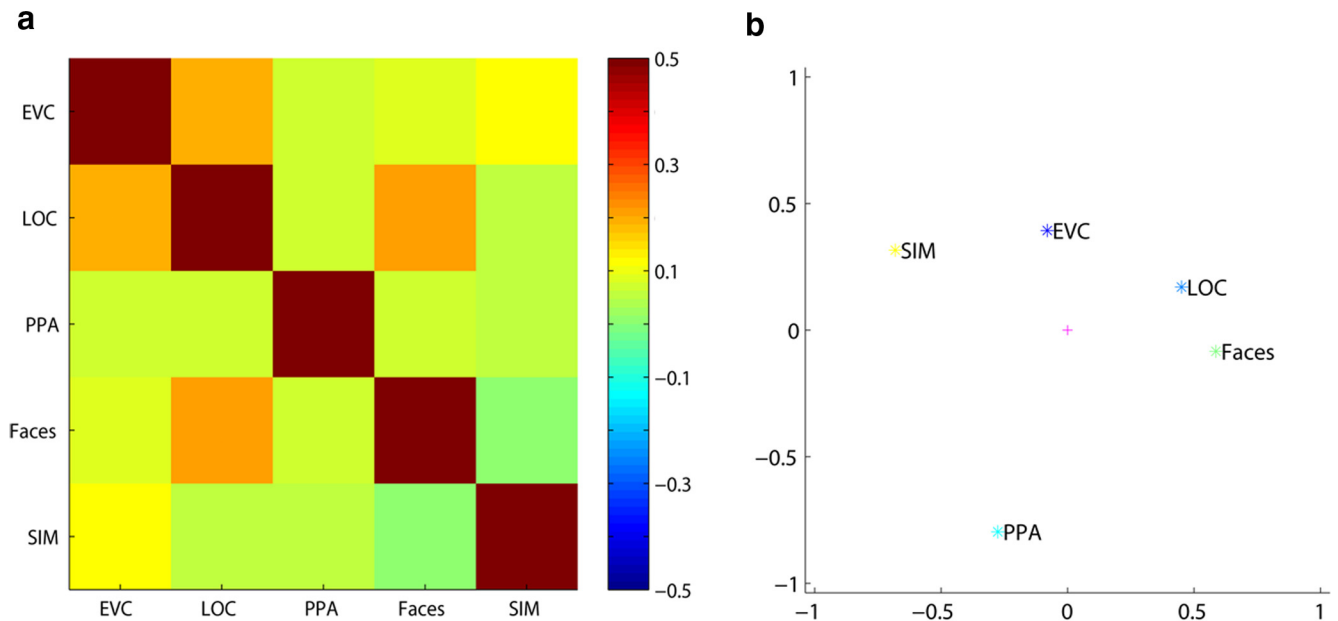
a

b

Figure 10. Overview of the correlation analyses. **a**, Matrix representing the correlations of the SVM performance matrices between all regions (EVC, LOC, PPA, and the joined face region "Faces"), as well as the pixel-based differences between the nine stimuli (SIM). **b**, Visual presentation of the results of principal component analysis applied to the correlation matrix shown in **a**. The physical differences between stimuli are more closely related to EVC than to the other regions and LOC falls in between EVC and the face regions.

was performed to look at configural and featural information before and after training. No significant effect of training ($F_{(1,11)} = 1.43$, $p = 0.26$), no difference between the two types of perceptual information ($F_{(1,11)} = 2.61$, $p = 0.14$), and no interaction between the two ($F < 1$, ns) were found. Again, decoding ability for configural and featural information was taken together as one category of perceptual information. Across sessions, there was significant decoding for perceptual information (sensitivity = 0.23, SEM = 0.060, 2-tailed paired t test for the difference from a sensitivity of zero, averaged across sessions: $t_{(11)} = 3.77$, $p = 0.0031$). In addition, decoding ability of perceptual information was significant both before (2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = 2.32$, $p = 0.041$) and after training ($t_{(11)} = 3.48$, $p = 0.0052$; Fig. 9b). In EVC, there was no representation of the scene associated with faces during training (posttraining decoding of the associated scene, 2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = 1.16$, $p = 0.27$; Fig. 9b).

Before training (2-tailed paired t test for the difference from a sensitivity of zero: $t_{(11)} = 2.28$, $p = 0.044$) and after training ($t_{(11)} = 6.73$, $p = 0.000033$), there was decoding for the difference between the base face and the other faces. Unlike LOC, this effect between sessions did differ significantly ($t_{(11)} = -5.39$, $p = 0.00022$; Fig. 9b).

Differences in representations across regions
Up to this point, we have grouped face pairs, for example, into configural versus featural changes. However, for each ROI, we have a full matrix with SVM performances for all of the possible face pairs. To investigate further the similarities and differences between the ROIs, we compared the matrices of the different ROIs averaged across pretraining and posttraining datasets: joined face region, PPA, EVC, and LOC. We did not further differentiate between the different face regions because statistical analyses did not reveal any differences among them. As a null model, we hypothesized that the SVM performance might be explainable by the physical differences among face images. We

calculated these physical differences by subtracting the corresponding pixel values of two images squaring the pixel values, summing across pixels, and taking the square root of this sum. These pixel-based difference scores were correlated with the SVM performance matrices of the different ROIs. We used the matrices for individual subjects to assess the significance of the correlations with a t test across subjects. Only in EVC did the SVM performance correlate with pixel-based image differences (average individual-subject correlation $r = 0.13$, $t_{(11)} = 2.7$, $p = 0.021$; p-values of the other regions > 0.22). Correlations between regions were also observed, in particular between LOC and EVC ($r = 0.19$, $t_{(11)} = 4.69$, $p = 0.00067$, also significant after FDR correction) and between LOC and the joined face area ($r = 0.21$, $t_{(11)} = 3.12$, $p = 0.0098$, also significant after FDR correction). Figure 10 shows the results of the correlation analysis (Fig. 10a) and a visual representation of the similarities and differences between these regions after performing a principal component analysis (Fig. 10b). These results seem to show a progression from purely physical differences that are more closely correlated with EVC to a possibly more complex representation in the regions situated more anterior in the ventral visual stream. The nature of these different representations remains to be investigated.

Discussion
In this study, a face set was generated to investigate the neural representations in the face regions of the ventral visual stream and to investigate whether perceptual and/or context information about faces could be successfully decoded using MVPA. This is the first study to systematically manipulate differences in both perceptual and contextual elements of face identity using MVPA.

Representation of perceptual differences among faces
Decoding of face identity information and information about the perceptual characteristics of faces, whether they are featural changes or changes in spacing between face parts, is present when considering the three face regions as a whole. When looking at a

more specific level, the three face regions do not show a distinction between the two types of perceptual information. When configural and featural information is considered together, perceptual information seems to be present in all three regions.

This representation of the perceptual properties of faces was by no means restricted to the face regions. To the contrary, it was present in all ROIs in the visual cortex. Similarity analyses indicated that not all of these visual regions represent faces in the same way. In EVC, the representations were most closely associated with pixel-based differences among face images, and this physical representation transformed from EVC over object-selective cortex to the face regions. This finding of progression is consistent with how visual stimuli in general are coded in the hierarchical visual system and extends previous findings with natural stimuli (Kriegeskorte et al., 2008) and artificial object shapes (Op de Beeck et al., 2008).

More to our surprise, the scene-selective area PPA also represented the perceptual differences among faces, and it was the only ROI with a specific preference for one perceptual dimension, with more sensitivity for configural than for featural differences. Because PPA is a region that responds more strongly to scenes (Epstein and Kanwisher, 1998) or, as suggested by some, contexts (Bar and Aminoff, 2003; Bar, 2004), this seems a counterintuitive finding. However, most studies that looked into the role of PPA found that it seems to be engaged predominantly during the processing of spatial relationships in scenes (Epstein, 2008; Epstein and Ward, 2010; Kravitz et al., 2011). Because configural processing concerns the spatial relations between face parts, the configuration-unique sensitivity in PPA might reflect spatial processing that is sufficiently generic to apply to faces (and objects) as much as to scenes. More research is needed to corroborate this intriguing hypothesis.

Representation of the contextual associations of faces
Which regions in the brain represent the context that is associated with a face? The findings in the face regions are clear-cut: no effects related to contextual information could be detected on an individual basis. This conclusion stands in apparent contrast with a recent study. Van den Hurk et al. (2011) were able to make a distinction between words grouped to represent information linked to different faces. However, given that the participants in that study were instructed to actively associate each presented word with the associated face image and given that the activity of category-selective regions generalizes from perception to imagery (Reddy et al., 2010), it is possible that the findings of Van den Hurk et al. (2011) could be attributed to the fact that participants were imagining the face referred to by the words.

In PPA, we reported a clear training-induced representation of the scene associated with a face. These results are consistent with the view that PPA is mainly concerned with scene processing and spatial relations between scenes (e.g., subjects might have imagined the associated scene during the last scan session), as well as with the view that PPA represents the context associated with an object (Bar and Aminoff, 2003). Given that we only measured the sensitivity to the associated scene and not to other associations (e.g., name, type of interaction with the person, etc.), our findings do not reveal whether PPA only represents the spatial and scene context of a face or if it has a more general function in the representation of context information.

Face familiarity
In most studies in the literature (Sugiura et al., 2001; Gobbini et al., 2004; Eifuku et al., 2011), the effects of familiarity were tested

by comparing responses to unknown faces with responses to famous faces or faces of people known to subjects. In such a comparison, there are many differences between the two conditions: the familiar faces have been seen more often and have many associations. Our training protocol also induces all these differences between the trained/familiar faces and the "unfamiliar" base face. Previous studies (Sergent et al., 1992; Gorno-Tempini et al., 1998; Leveroni et al., 2000) suggested that fusiform and anterior temporal regions respond more strongly to familiar than to unfamiliar faces. In our study, there was no similar effect of training on the mean response strength to both familiar and unfamiliar faces, but we observed an analogous finding with respect to the spatial response patterns when considering the face regions as a whole: familiar and unfamiliar faces were both associated with a different response pattern after the training had induced a difference in familiarity. In addition, other regions showed this pattern as well: we found similar effects in EVC and a trend in LOC. Because our context manipulations contained many factors related to familiarity of the stimuli, it is perhaps not that surprising that a number of different regions seem to be affected by these differences. Nevertheless, the investigation of effects of familiarity was not the primary aim of our study, and as such it was not designed to allow firm conclusions about how and why the different face regions were affected by familiarity as defined in our study.

Sensitivity to face information can be found in LOC and EVC
Averaged across sessions, LOC was able to differentiate between different faces. Furthermore, LOC is sensitive to all perceptual differences among faces (featural and configural) and to the distinction between the untrained base face and the trained faces. Clearly, differences among faces are not only represented in face regions. We also found a significant decoding of the differences among faces in EVC. This finding is consistent with the observation by Mur et al. (2010), who found adaptation effects in non-face regions, including PPA and EVC.

It is not because all of these regions represent differences among the faces in our stimulus set that all of these representations are the same. In our experiment, we had already found that only the EVC representation was related to pixel-based differences between faces and that the representation in the joined face region was not related to the representation in EVC. This pattern of results suggests the presence of a higher-level representation in the face regions. Studies including manipulations designed to dissociate low- and high-level representations, such as viewpoint changes, would probably show even stronger differences between the representations in EVC, LOC, and face regions.

Overall, our findings support the notion that the information about faces is distributed widely across the visual system. Perceptual dimensions are coded across all visual regions, with a shift in the nature of this representation from EVC to face regions. Contextual information about locations associated with faces is not represented in face regions per se, but in other regions, notably the PPA.

References
Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617–629. CrossRef Medline
Bar M, Aminoff E (2003) Cortical analysis of visual context. Neuron 38: 347–358. CrossRef Medline
Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436. CrossRef Medline
Davidenko N, Remus DA, Grill-Spector K (2012) Face-likeness and image

variability drive responses in human face-selective ventral regions. Hum Brain Mapp 33:2334–2349. CrossRef Medline

Eifuku S, De Souza WC, Nakata R, Ono T, Tamura R (2011) Neural representations of personally familiar and unfamiliar faces in the anterior inferior temporal cortex of monkeys. PLoS One 6:e18913. CrossRef Medline

Epstein RA (2008) Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn Sci 12:388–396. CrossRef Medline

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598–601. CrossRef Medline

Epstein RA, Ward EJ (2010) How reliable are visual context effects in the parahippocampal place area? Cereb Cortex 20:294–303. CrossRef Medline

Gobbini MI, Leibenluft E, Santiago N, Haxby JV (2004) Social and emotional attachment in the neural representation of faces. Neuroimage 22:1628–1635. CrossRef Medline

Gorno-Tempini ML, Price CJ, Josephs O, Vandenberghe R, Cappa SF, Kapur N, Frackowiak RS, Tempini ML (1998) The neural systems sustaining face and proper-name processing. Brain 121:2103–2118. CrossRef Medline

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430. CrossRef Medline

Kanwisher N, Yovel G (2006) The fusiform face area: a cortical region specialized for the perception of faces. Philos Trans R Soc Lond B Biol Sci 361:2109–2128. CrossRef Medline

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311. Medline

Kravitz DJ, Peng CS, Baker CI (2011) Real-world scene representations in high-level visual cortex: it's the spaces more than the places. J Neurosci 31:7322–7333. CrossRef Medline

Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci U S A 104:20600–20605. CrossRef Medline

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis- connecting the branches of systems neuroscience. Front Syst Neurosci 2:4. CrossRef Medline

Leopold DA, Bondar IV, Giese MA (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442:572–575. CrossRef Medline

Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM (2000) Neural systems underlying the recognition of familiar and newly learned faces. J Neurosci 20:878–886. Medline

Loffler G, Yourganov G, Wilkinson F, Wilson HR (2005) fMRI evidence for the neural representation of faces. Nat Neurosci 8:1386–1390. CrossRef Medline

Mur M, Ruff DA, Bodurka J, Bandettini PA, Kriegeskorte N (2010) Face-

identity change activation outside the face system: "release from adaptation" may not always indicate neuronal selectivity. Cereb Cortex 20:2027–2042. CrossRef Medline

Nestor A, Vettel JM, Tarr MJ (2008) Task-specific codes for face recognition: how they shape the neural representation of features for detection and individuation. PLoS One 3:e3978. CrossRef Medline

Nestor A, Plaut DC, Behrmann M (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. Proc Natl Acad Sci U S A 108:9998–10003. CrossRef Medline

Op de Beeck HP, Torfs K, Wagemans J (2008) Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. J Neurosci 28:10111–10123. CrossRef Medline

Op de Beeck HP, Brants M, Baeck A, Wagemans J (2010) Distributed subordinate specificity for bodies, faces, and buildings in human ventral visual cortex. Neuroimage 49:3414–3425. CrossRef Medline

Pinsk MA, Arcaro M, Weiner KS, Kalkus JF, Inati SJ, Gross CG, Kastner S (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. J Neurophysiol 101:2581–2600. 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

Rajimehr R, Young JC, Tootell RB (2009) An anterior temporal face patch in human cortex, predicted by macaque maps. Proc Natl Acad Sci U S A 106:1995–2000. CrossRef Medline

Reddy L, Tsuchiya N, Serre T (2010) Reading the mind's eye: Decoding category information during mental imagery. Neuroimage 50:818–825. CrossRef Medline

Sergent J, Ohta S, MacDonald B (1992) Functional neuroanatomy of face and object processing–a positron emission tomography study. Brain 115:15–36. CrossRef Medline

Sugiura M, Kawashima R, Nakamura K, Sato N, Nakamura A, Kato T, Hatano K, Schormann T, Zilles K, Sato K, Ito K, Fukuda H (2001) Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances. Neuroimage 13:877–890. 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

van den Hurk J, Gentile F, Jansma BM (2011) What's behind a face: person context coding in fusiform face area as revealed by multivoxel pattern analysis. Cereb Cortex 21:2893–2899. CrossRef Medline

Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. J Am Med Inform Assoc 8:443–459. CrossRef Medline

Weiner KS, Grill-Spector K (2010) Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. Neuroimage 52:1559–1573. CrossRef Medline