Behavioral/Cognitive

Single-Unit Recordings in the Macaque Face Patch System Reveal Limitations of fMRI MVPA

Julien Dubois,[1] Archy Otto de Berker,[2] and Doris Ying Tsao[1]
[1]Division of Biology, California Institute of Technology, Pasadena, California 91125, and [2]Pembroke College, University of Cambridge, Cambridge CB21RF, United Kingdom

Multivariate pattern analysis (MVPA) of fMRI data has become an important technique for cognitive neuroscientists in recent years; however, the relationship between fMRI MVPA and the underlying neural population activity remains unexamined. Here, we performed MVPA of fMRI data and single-unit data in the same species, the macaque monkey. Facial recognition in the macaque is subserved by a well characterized system of cortical patches, which provided the test bed for our comparison. We showed that neural population information about face viewpoint was readily accessible with fMRI MVPA from all face patches, in agreement with single-unit data. Information about face identity, although it was very strongly represented in the populations of units of the anterior face patches, could not be retrieved from the same data. The discrepancy was especially striking in patch AL, where neurons encode both the identity and viewpoint of human faces. From an analysis of the characteristics of the neural representations for viewpoint and identity, we conclude that fMRI MVPA cannot decode information contained in the weakly clustered neuronal responses responsible for coding the identity of human faces in the macaque brain. Although further studies are needed to elucidate the relationship between information decodable from fMRI multivoxel patterns versus single-unit populations for other variables in other brain regions, our result has important implications for the interpretation of negative findings in fMRI multivoxel pattern analyses.

Key words: face identity; face patch system; face viewpoint; fMRI MVPA; macaque; single-unit populations

Introduction

The ability of fMRI multivariate pattern analysis (MVPA) to infer the orientation of visual gratings from the pattern of BOLD activity in the human primary visual cortex (V1) (Haynes and Rees, 2005; Kamitani and Tong, 2005) established the technique as an important counterpart to classical univariate analyses. Although the relationship between underlying neural population activity and fMRI MVPA is still currently unclear, fMRI decoding has been applied in many areas of cognitive neuroscience, encompassing perception, attention, object processing, memory, semantics, language processing, and decision making (for review, see Tong and Pratte, 2012). The interpretation of above-chance classification in terms of brain function, however, critically rests on a yet-to-be-established link between neural population activity and fMRI patterns.

A fundamental step toward understanding the neural basis of MVPA is the comparison of information encoded by populations of single units with that of information decoded from fMRI patterns. The macaque face patch system (Tsao et al., 2003, 2008) offers an unprecedented opportunity for such an investigation. Highly reproducible across animals (although there remains debate about the precise number of patches and their nomenclature; Pinsk et al., 2009; Ku et al., 2011; Rajimehr et al., 2014; Vanduffel et al., 2014) and exquisitely functionally compartmentalized (Freiwald and Tsao, 2010), the face patch system permits a direct comparison of the two techniques in the same regions and in the same species. Notably, the selectivity of single neurons for face viewpoint and face identity in a subset of the face patches (the middle face patches ML and MF, the anterior lateral face patches AL and AF, and the anterior medial face patch AM) differs greatly. For example, neurons become increasingly view invariant as one moves anteriorly, with the emergence of mirror-symmetric tuning in AL and fully view-invariant tuning in the most anterior patch, AM. Here, we apply multivariate analysis to single-unit recordings and to fMRI data in these patches and compare information retrieved by linear classifiers from data collected by the two recording methods.

Beyond shedding light on the neural underpinnings of fMRI decoding, our results also inform the literature on MVPA studies of face identity in the human brain (Kriegeskorte et al., 2007; Natu et al., 2010; Nestor et al., 2011; Anzellotti et al., 2014). Some have claimed that the fusiform face area contains identity information (Nestor et al., 2011; Anzellotti et al., 2014), whereas others failed to find such evidence (Kriegeskorte et al., 2007). Therefore, fMRI decoding has not yet brought a definite answer

to the question (fMRI adaptation studies have similarly failed to reach a consensus on the matter; Mur et al., 2010). We find that brain regions containing identity-specific neurons do not support decoding of facial identity with fMRI MVPA; further analysis of the population code demonstrates that this failure may be due to weak clustering of like-tuned units. Conversely, readout of spatially clustered representations of facial viewpoint shows striking agreement between fMRI MVPA and single-unit recordings.

Materials and Methods

Procedures. All procedures conformed to local and National Institutes of Health guidelines, including the National Institutes of Health *Guide for Care and Use of Laboratory Animals.* All experiments were performed with the approval of the Institutional Animal Care and Use Committee.

Stimuli. Freiwald and Tsao (2010) used a set of 200 photographs of human faces comprising 25 different identities, each taken from eight different viewpoints, which they refer to as the face views (FV) image set. We randomly picked five males of the 25 identities and selected five of the eight viewpoints: left full profile (L90), left half profile (L45), frontal (F), right half profile (R45), and right full profile (R90); this left us with a set of 25 images (Fig. 1a). This set of images was used for the fMRI experiments.

Single-unit data acquisition and experimental paradigm. Most of the single-unit data came from an existing dataset and the reader is referred to Freiwald and Tsao (2010) for a full description of these recordings. Data were recorded in three male rhesus macaques. Face patches were localized in each animal using fMRI. The animals were implanted with Ultem headposts and the following face patches were targeted for single-unit recordings with a fine tungsten electrode: MF, AL and AM for M1; AL and AM for M2; and ML for M3. The monkeys were rewarded with juice for constant fixation while viewing all pictures from the FV image set (200 pictures total) in random order without replacement; depending on the cell, the set of 200 images was shown between 3 and 10 times. Each image was shown for 200 ms with a 200 ms blank between images. Only well isolated single units that showed a refractory period were studied. Data were collected over multiple sessions (M1, MF: 7 sessions, AL: 12 sessions, AM: 23 sessions; M2, AL: 59 sessions, AM: 13 sessions; M3, ML: 26 sessions).

We also report new data from electrophysiological recordings in AM for monkey M5. M5 was presented with the FV image set; depending on the cell, each image was shown between 1 and 17 times. Each image was shown for 150 ms with a 150 ms blank between images. Data were collected over 20 sessions. In addition, we collected the responses of 6 AM cells to 24-s-long presentations of the frontal views for the 5 identities used in the fMRI experiments, jittering the images every 2 s to avoid retinal adaptation (see fMRI paradigm).

Single-unit decoding. As argued in Freiwald and Tsao (2010), the similarity of responses obtained from the same patch in different animals warranted pooling data across animals. Furthermore, the similarity of responses in ML and MF warranted pooling data from these patches and we refer to this merged patch as ML/MF. We selected all units that had been presented with all 25 images in our set a minimum of three times.
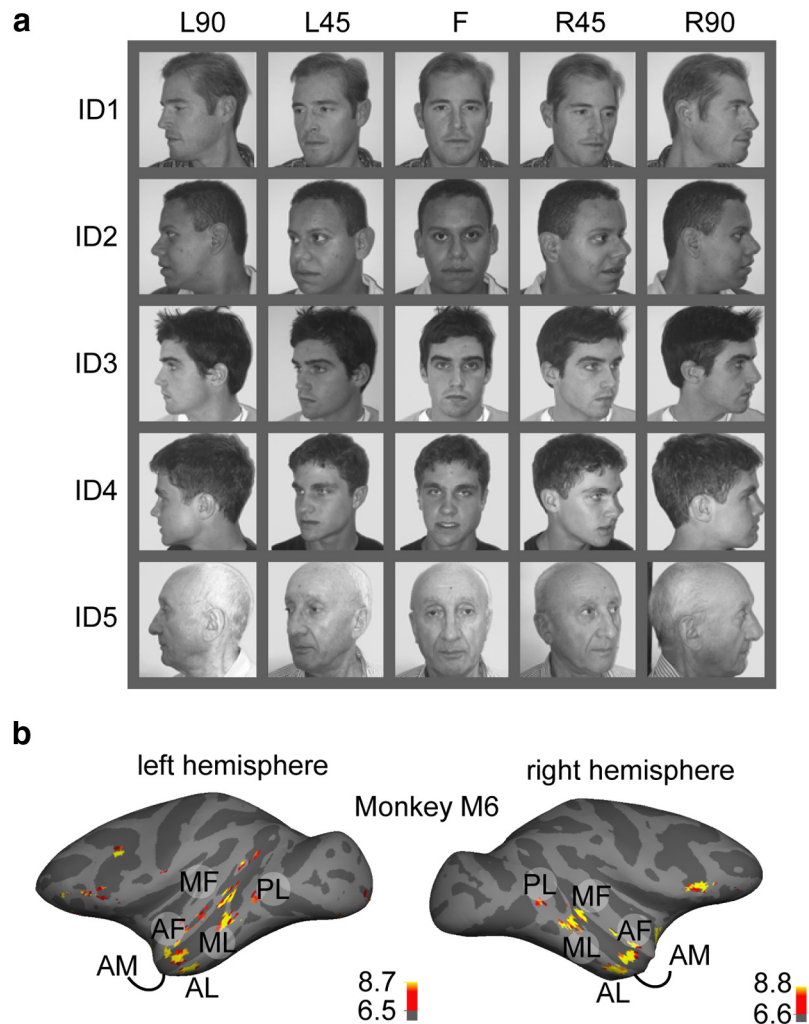
Figure 1. Face stimuli for the experiments and brain regions of interest (face patches). *a*, The 25 images used in the single-unit recordings and fMRI experiments comprising five male identities each depicted from five viewpoints. *b*, Statistical parametric map for the contrast faces versus {bodies, fruits, hands, and objects} for monkey M6 displayed on the inflated left and right hemispheres; the main clusters, corresponding to a subset of the face patches (PL, ML/MF, AL/AF and AM), are labeled. The color scale represents T-values.

This criterion yielded a total of 66 units in ML/MF, 102 units in AL, and 167 units in AM. The response of each unit to each image was defined as the firing rate in the 50–200 ms window, expressed as a percentage of the baseline firing rate in the 0–50 ms window. We randomly selected three trials (of the three to 10 available trials) for each of the 25 images (we repeated this random selection procedure 20 times to have a better estimator of the true decoding performance) and, for each patch, we built a matrix with 75 rows representing examples (25 images × three presentations) and as many columns as there were neurons available. We used a threefold cross validation scheme. At each fold, we set aside one of the three trials for each image to serve as the testing set and trained the classifier on the remaining two trials; we repeated the threefold cross-validation procedure 10 times to achieve a more robust analysis. To test for identity-invariant viewpoint information, we restricted the testing examples to only one of the five identities (we did this for each identity in turn) and the training examples to the remaining four identities. Similarly, to identify viewpoint-invariant identity information, we trained the classifier on four of the five viewpoints and tested on the remaining viewpoint. We normalized the training and testing sets by removing the column mean of the training set and dividing by the column SD of the training set. Then, we trained a linear Support Vector Machine (LIBSVM for MATLAB downloaded from http://www.csie.ntu.edu.tw/~cjlin/libsvm) to discriminate either the viewpoint or the identity in the train-

ing set (note that we did not optimize the C parameter, which regulates the trade-off between misclassifications and margin, because it did not make any difference within a reasonable range; we used the default C = 1). Finally, we applied the learned classifier on the testing set. We used the "−b1" option in LIBSVM, which computes probability estimates for each class (through an internal fivefold cross validation; see Wu et al., 2004) and can yield predictions that differ slightly from the "−b0" nonprobabilistic option; however, it also provides a better picture of the information available to the classifier by recording how difficult each decision is (which a simple confusion matrix does not do). Confusion matrices were populated by counting the predicted labels of each class type for each input class; the overall accuracy was computed as the sum of the diagonal elements of the confusion matrix divided by the sum of all elements. Note that linear SVM does not support multiclass classification; instead, multiclass problems need to be reformulated as a series of binary classifications. LIBSVM internally uses an "all vs all" scheme (sometimes referred to as "one vs one"), whereby as many binary classifications are run as there are pairs of labels.

fMRI data acquisition and experimental paradigm. Five male rhesus macaques (M4, M5, M6, M7, and M8) were trained to maintain fixation on a small spot for a juice reward. Monkeys were scanned in a 3T TIM (Siemens) scanner while passively viewing images on a screen. Eye position was monitored using an infrared eye tracking system (ISCAN) and a juice reward was delivered every 2–4 s if fixation was properly maintained. The fixation spot size was 0.13° in diameter. We used a gradient-echo EPI sequence (TR 2 s, TE 17 ms, flip angle 80°, 96×96 resolution, 54 slices, 1 mm isotropic resolution, parallel imaging GRAPPA with acceleration factor 2, phase partial Fourier 7/8) for M4, M5, M6, and M7 and a slightly different sequence (EPI, TR 2 s, TE 17 ms, 80×80 resolution, flip angle 80°, 45 slices, 1.2 mm isotropic resolution, no parallel imaging, phase partial Fourier 6/8) for M8. In combination with a concomitantly acquired field map, this allowed high-fidelity reconstruction by undistorting most of the B0-field inhomogeneities (Zeng and Constable, 2002; Cusack et al., 2003). MION contrast agent was used to improve signal-to-noise ratio (SNR). Images presented on the screen spanned 9.4° of visual angle. Twenty-four-second blocks of a gray background alternated with 24 s blocks of the images. The same image (1 of 25) was presented throughout an image block, with its position jittered slightly (0.9°) every 2 s to prevent visual adaptation. We collected 10 fMRI runs for M4 (one session), 10 runs for M5 (one session), 14 runs for M6 (two sessions), 38 runs for M7 (five sessions), and 14 runs for M8 (two sessions). For M4 and M5, during each run, we presented 10 images; therefore, it took two runs to present all 20 images (we did not present ID4 to M4 and M5). The order of images was fixed, and balanced (run A: ID2, F; ID1, L90; ID5, F; ID3, F; ID2, L45; ID5, L45; ID3, R45; ID2, L90; ID1, R90; ID5, R45; run B: ID3, R90; ID1, F; ID1, L45; ID1, R45; ID5, L90; ID3, L90; ID2, R45; ID5, R90; ID2, R90; ID3, L45). For M6, M7, and M8, we presented either 12 or 13 images in each run (again, it took two runs to present all 25 images). The order of images was pseudorandomized independently for each pair of runs, with the constraint that all identities and viewpoints were presented in each run.

fMRI preprocessing. EPI data were realigned to the first run and corrected for distortions caused by magnetic field inhomogeneities using Freesurfer (downloaded form http://surfer.nmr.mgh.harvard.edu/). The short TR (2 s) did not warrant the application of slice timing correction. All further analyses were performed in MATLAB.

Face patch localization. We acquired data to functionally localize face patches in a separate fMRI session for each monkey. Five face-selective regions (ML, MF, AL, AF, and AM) were identified in each hemisphere in all monkeys using a univariate contrast between blocks of faces (monkey faces and human faces, familiar and unfamiliar) versus other categories (bodies, fruits, hands, and man-made objects). Additional details have been described previously (Tsao et al., 2006; Freiwald and Tsao, 2010; Ohayon and Tsao, 2012). We thresholded the statistical parametric maps at $p < 0.0001$ uncorrected and selected clusters of contiguous voxels to define the face patches (we masked each face patch with a 1-cm-diameter sphere centered on the peak voxel of each cluster).

fMRI decoding. We first detrended the time course of each voxel in each run independently using a second-order polynomial, then z-scored

the signal across time (Kietzmann et al., 2012). We took the average of time points 16, 18, 20, 22, 24, and 26 s (which encompass the peak of the fMRI response to the block stimulation) as the signal for each block. We extracted the signal for each block at each voxel ($n_{vox} = 96 \times 96 \times 54$ for M4–M7, $80 \times 80 \times 45$ for M8), thus populating a ($n_{blocks} \times n_{vox}$) matrix. We then selected the columns of this matrix that corresponded to each functionally defined region of interest (face patches) and submitted the data from each ROI to a multivariate pattern analysis. A commonly used machine learning algorithm for supervised pattern classification in the fMRI decoding literature is the linear Support Vector Machine (which we also used for decoding from single-unit pseudopopulations). Other common choices include Linear Discriminant Analysis, Gaussian Naive Bayes, or sparse logistic regression; we chose to use linear SVM because it usually performs at least as well as any other classifier on fMRI data (Pereira and Botvinick, 2011). The procedure was similar to that described in the single-unit decoding section except that there was no random selection of data; we used a leave-one-run-out cross-validation scheme, thus avoiding any dependence between the testing examples and training examples. Beyond the analyses reported in the main text, we explored additional analyses to convince ourselves that we could not perform better with these approaches. These included, for example, smoothing the data with a Gaussian kernel before decoding and/or expanding the regions of interest and then using feature selection approaches. None of these analyses yielded significant decoding in regions where we did not find significant decoding with the classical approach.

Searchlight decoding. We used a cubic searchlight comprising 125 voxels ($5 \times 5 \times 5$) and ran it throughout the fMRI volume; within each searchlight, we used a linear SVM classifier in the exact same way as we did in the ROI analyses described in the previous section. Average accuracy was mapped to the voxel at the center of the searchlight.

Statistical analysis of decoding performance. Statistical assessment of decoding performance is an area of ongoing debate (Schreiber and Krekelberg, 2013; Noirhomme et al., 2014); current best practice is to perform a permutation test whereby one assigns wrong labels to training examples and conducts the whole analysis (including scaling, feature selection, etc.) using these surrogate labels. Unless specified otherwise, we report the 95% interval for 1000 surrogate datasets as a vertical line (which should be centered on chance level) in the majority of figures. We derived all p-values from these permutation tests by counting how many of the surrogate results are equal to or better than the real result and dividing by the number of surrogates. Note that, for searchlight decoding analyses, it was computationally too expensive to perform permutation testing so we had to revert to binomial testing.

Representational similarity. Our invariant decoding procedure generated five probability matrices, one for each of the testing identities (resp. viewpoints). We concatenated all elements of these five matrices into a single vector, yielding a detailed description of the information available to the classifiers. We then computed the Spearman rank correlation (ρ) between the vectors corresponding to the single unit decoding and the vectors corresponding to the fMRI decoding. We assessed the significance of these correlations with a permutation test, drawing 1000 combinations of surrogates from the single-unit and fMRI data. We also computed these correlations considering only nondiagonal elements of each probability matrix to focus on the pattern of errors.

Tuning analyses. To investigate the tuning of single neurons and of single voxels to viewpoint and identity, we used the method described previously (Serences et al., 2009; Gratton et al., 2013). Here, we describe the procedure for the single voxels and with viewpoint (the procedure is similar for single units) and with identity. For each trial, we computed the mean response across all voxels in a given patch and removed it from each voxel's response to correct for mean effects and then z-scored the responses for each voxel across trials in each fMRI run. We next assigned each voxel a preferred viewpoint based on its normalized response to each viewpoint averaged over all but one run (training runs); the preferred viewpoint was the one evoking the largest mean normalized response in the training runs. At the end of this process, voxels were sorted into five classes according to their viewpoint preference in the training runs. We finally computed the mean normalized response of all voxels in each class to each of the five viewpoints in the testing run, resulting in a

tuning function for each class. This procedure was repeated leaving each run out in turn (cross-validation), and tuning functions were averaged across folds. To characterize the amount of tuning to the preferred viewpoint and compare it between single voxels and single units, we z-scored each final tuning curve and computed the difference between the normalized response to the preferred viewpoint and the average normalized response to all other viewpoints: we name the resulting quantity the "tuning factor," expressed in units of SDs from the average response. We computed an average tuning factor for each face patch from its five tuning curves.

We also computed a mutual information based measure to quantify the tuning of each voxel to viewpoint. Normalized responses in the training runs were discretized into five bins based on the range of responses across all voxels. We computed the entropy of the binned responses, $H(B)$, for each voxel as follows:

$$H(B) = -\sum_{b \varepsilon B} p(b) log_2 p(b), \qquad (1)$$

where $p(b)$ is the proportion of trials in which the voxel's response falls into bin b. Then, we computed the conditional entropy $H(B|VP)$, the entropy of the responses for each voxel given knowledge of the viewpoint, as follows:

$$H(B|VP) = -\sum_{vp \varepsilon VP} p(vp) \sum_{b \varepsilon B} p(b|vp) log_2 p(b|vp), \qquad (2)$$

From these two quantities, the mutual information $MI(B;VP)$, the viewpoint information carried by the responses of a voxel, is therefore as follows:

$$MI(B;VP) = H(B) - H(B|VP). \qquad (3)$$

The unit of measure is bits (base 2 logarithm). Informative voxels have a high MI. Statistical assessment of the tuning analyses was conducted through permutation testing.

Sparseness. We computed the Gini index (Hurley and Rickard, 2009) on the basis of the normalized average responses to each identity (and to each viewpoint) in the face views set for all neurons in a given patch. The normalized responses represent how strongly each neuron responds to each identity (or viewpoint) as a fraction of their maximal response (in the image set); the response is sparse if a given stimulus evokes a maximal response in only a few neurons.

$$G = 1 - 2\sum_{k=1}^{N} \frac{c_{(k)}}{\|c\|} \left(\frac{N - k + \frac{1}{2}}{N} \right) for\ ordered\ data,$$

$$c_{(1)} \le c_{(2)} \le \dots \le c_{(N)} \qquad (4)$$

Clustering. For the clustering analysis, we pairwise correlated the average patterns of responses (averaged by identity or by viewpoint, respectively) of units recorded in the same penetration, within 1 mm of each other. The numbers of pairs satisfying this criterion was as follows: ML/MF, 385; AL, 326; AM, 610. For each patch, we computed the average correlation score (after a Fisher Z transform) across all these pairs. We used a permutation test to assess the statistical significance of the resulting average correlations. We randomly shuffled the identity and viewpoint labels of the data and computed the pairwise correlation between all neurons satisfying our distance criterion. We repeated this 1000 times to compose a null distribution against which we tested our observed correlations.

Results

Single-unit tuning to face viewpoint and identity in the face patches

Most of the single-unit data came from an existing dataset and the reader is referred to Freiwald and Tsao (2010) for a full description of these recordings performed in monkeys M1, M2, and M3. We randomly picked five human (male) identities from the image set of 25 identities described in Freiwald and Tsao (2010) and

selected five of the eight viewpoints in that set (left full profile L90, left half profile L45, frontal F, right half profile R45, and right full profile R90; leaving out up, down and back), thus yielding a total of 25 images to be used for the planned fMRI experiments (Fig. 1a).

The representations of viewpoint and identity at the level of single units in the face patches were investigated in detail in Freiwald and Tsao (2010). Here, we further characterized the tuning of single units to face viewpoint and identity using the methods described previously (Serences et al., 2009; Gratton et al., 2013): we established tuning functions and mutual information measures for our single-unit data for viewpoint and identity using a cross-validation procedure, as described in the Material and Methods section. Results for these analyses are reported in Figure 2 (top) for viewpoint and Figure 3 (top) for identity. In Figure 2 (top), empty bars in the left panel represent the proportion of neurons that have a higher response to each of the five viewpoints (in the training data); for example, in ML/MF, there are more neurons than expected by chance that have a higher response to the frontal view in the training set ($p < 0.01$) and fewer than expected by chance that have a higher response to the right profiles (R45, $p < 0.001$; R90, $p < 0.001$). The right panel shows the average response of the neurons in each class to each of the five viewpoints (in the testing data). Keeping with the example of the neurons tuned to the frontal view (in the training data), their average response in the testing data is higher than expected by chance for the frontal view ($p < 0.001$) and also lower than expected by chance for the R90 view ($p < 0.01$). Finally, the filled bars in the left panel represent the proportion of neurons that have a significant tuning to viewpoint according to the mutual information criterion in each class. There is a significant proportion of neurons tuned to four of the viewpoints (L90, $p < 0.001$; L45, $p < 0.001$; F, $p < 0.001$; and R90, $p < 0.05$). Summarizing results for viewpoint tuning, we observe the following. In ML/MF, there are single neurons significantly tuned to most viewpoints, with a predominance of frontal-view tuned units. The tuning curves present a single peak, with responses falling off on either side of the peak (sometimes asymmetrically). In AL, there is a significant proportion of neurons tuned to each of the viewpoints (L90, $p < 0.001$; L45, $p < 0.01$; F, $p < 0.001$; R45, $p < 0.01$; and R90, $p < 0.001$). Note that the tuning curves for profile-view tuned neurons are U-shaped: units that respond highly to the left (respectively right) full profile also respond highly to the right (respectively left) full profile. This is also apparent, although less pronounced, for half profile views. Neurons tuned to the frontal view respond little to profile views. In AM, there are only neurons significantly tuned to the frontal view according to the mutual information metric ($p < 0.001$). Note that the units classified as tuned to either left or right full profiles have a U-shaped tuning curve and respond less than expected by chance to the frontal view ($p < 0.05$).

Turning to identity tuning, our analyses indicate that overall there is very little tuning to identity in ML/MF neurons. Tuning curves in the testing set do not depart significantly from chance except for units tuned to ID3 that may respond slightly less than chance to ID5 ($p < 0.05$); note that there is a significant proportion of neurons tuned to ID3 ($p < 0.05$) according to the mutual information criterion. In AL, we find a significant proportion of units tuned to ID2 ($p < 0.05$), ID3 ($p < 0.001$), ID4 ($p < 0.01$), and ID5 ($p < 0.001$). The tuning curves show trends but do not significantly depart from chance. In AM, there are again significant proportions of neurons tuned to four of the five identities (ID2, $p < 0.001$; ID3, $p < 0.001$; ID4, $p < 0.05$; ID5, $p < 0.001$)
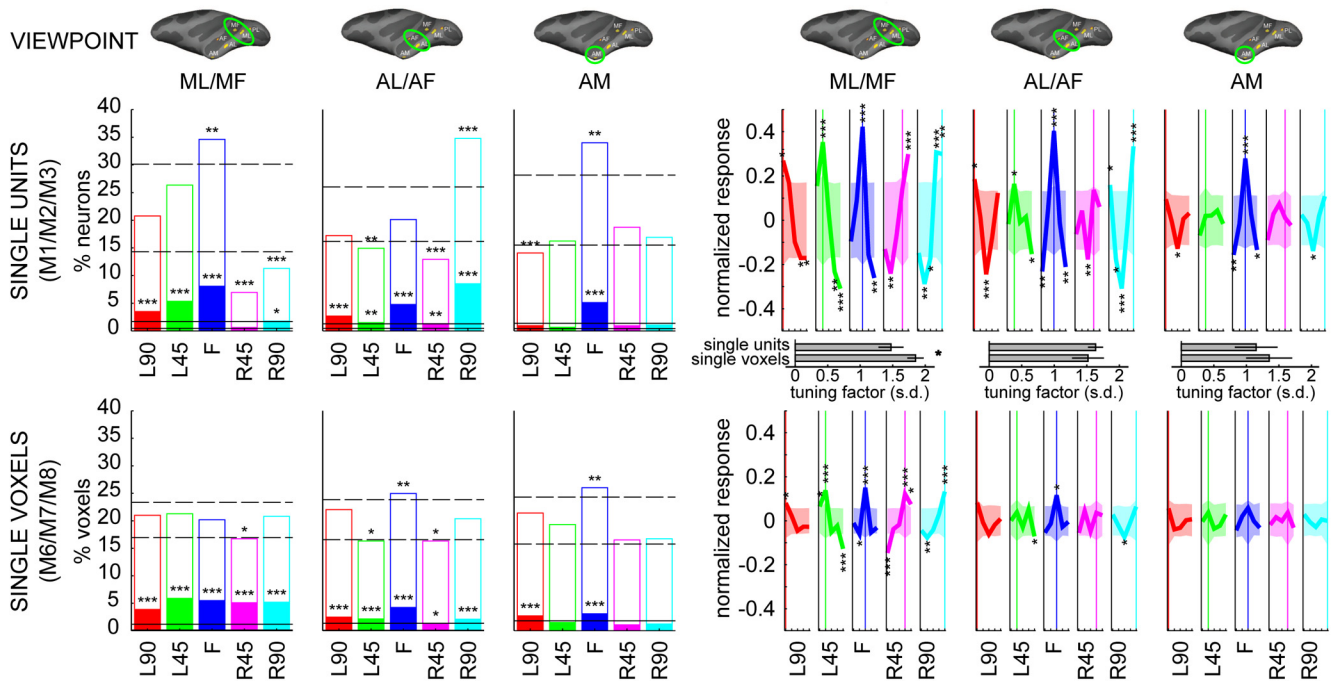
Figure 2. Single-unit and single-voxel tuning to viewpoint in the macaque face patches. Top, Single-unit tuning to viewpoint (based on all neurons recorded from M1, M2, and M3). In the left panel, the empty bars represent the proportions of neurons with the highest response for each of the five viewpoints (all bars sum to 100%). The filled bars represent the proportions of neurons in each class that are significantly tuned; that is, for which the mutual information (between neural activity and viewpoint) is higher than the 97.5th percentile of the null distribution. The right panel shows tuning curves for cells tuned to each of the five viewpoints within each patch. Null distributions, 95% confidence intervals (top, dashed black lines correspond to empty bars and solid black lines correspond to filled bars; bottom, shaded area) and p-values (*p < 0.05, **p < 0.01, ***p < 0.001) are all derived from a permutation test (1000 repetitions). Bottom, Single-voxel tuning to viewpoint (based on all voxels recorded from M6, M7, and M8). Middle, right, The tuning factor for each patch is compared between single units and single neurons; p-values are from a paired t test.
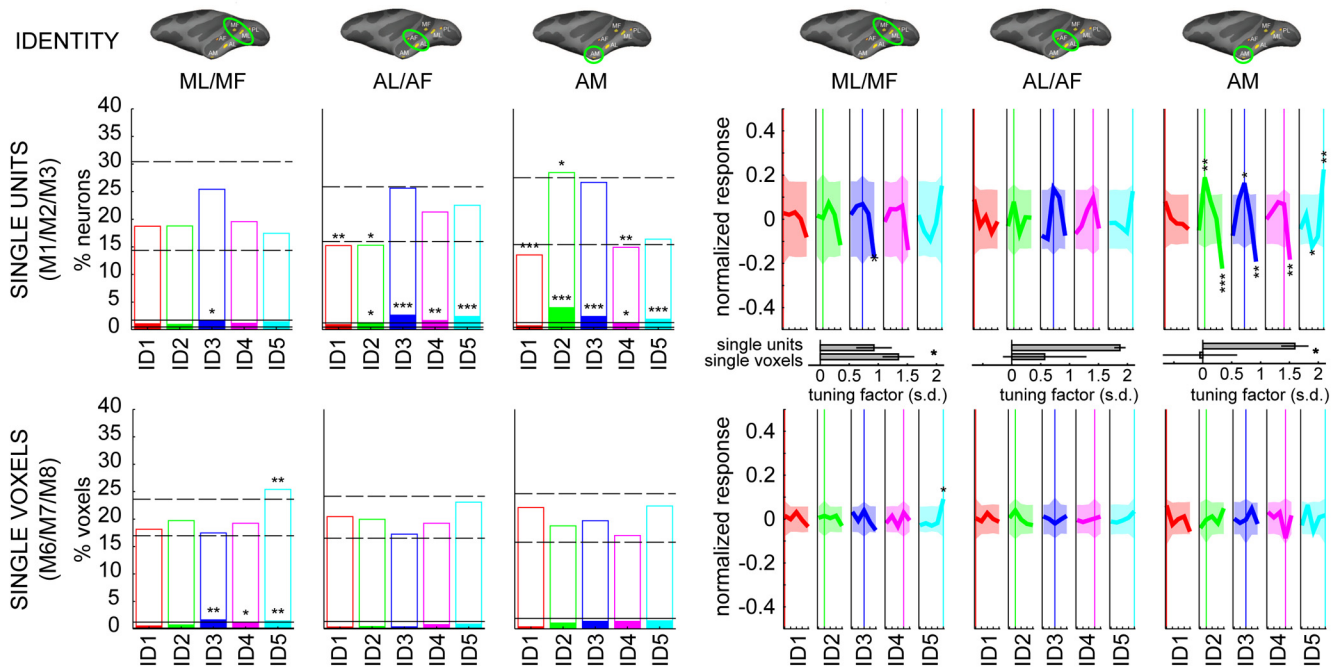

Figure 3. Single-unit and single-voxel tuning to identity in the macaque face patches. Top, Single-unit tuning to identity (based on all neurons recorded from M1, M2, and M3). In the left panel, the empty bars represent the proportions of neurons with the highest response for each of the five identities (all bars sum to 100%). The filled bars represent the proportions of neurons in each class that are significantly tuned; that is, for which the mutual information (between neural activity and identity) is higher than the 97.5th percentile of the null distribution. The right panel shows tuning curves for cells tuned to each of the five identities within each patch. Null distributions, 95% confidence intervals (top, dashed black lines correspond to empty bars and solid black lines correspond to filled bars; bottom, shaded area) and p-values (*p < 0.05, **p < 0.01, ***p < 0.001) are all derived from a permutation test (1000 repetitions). Bottom, Single-voxel tuning to identity (based on all voxels recorded from M6, M7, and M8). Middle, right, Tuning factor for each patch is compared between single units and single neurons; p-values are from a paired t test.

Table 1. Number of voxels in each of the regions of interest (early visual areas and face patches)

	V1	V2	V3	V4	ML-MF	AL-AF	AM
M4					259	178	52
M5					114	144	24
M6	2286	2141	1128	883	266	232	137
M7	1715	1736	1068	1135	481	308	152
M8*	1799	1572	1022	748	323	189	30

*Note that the resolution was 1.2 mm isotropic for M8 and 1 mm isotropic for all others.

and the tuning curves mostly reflect significant tuning as expected for each class (ID2, $p < 0.01$; ID3, $p < 0.05$; ID4, NS; ID5, $p < 0.01$).

Single-voxel tuning to viewpoint and identity in the face patches

We ran a block-design fMRI experiment in five monkeys (M4 through M8). In a separate fMRI session for each monkey, we ran a standard faces-objects-bodies localizer to functionally define the face patches (see Fig. 1b for the locations of these face patches on the inflated brain of M6; see Table 1 for the numbers of voxels in each face patch for each monkey). Here, we only present the results for M6, M7, and M8; although the data from M4 and M5 is entirely consistent with what we find in M6, M7, and M8, the experiments for M4 and M5 only included 4 identities and we had fewer data (10 runs, 1 session) for these 2 monkeys.

We were interested in establishing whether single voxels in the face patches are tuned to face viewpoint (respectively identity) using the same approach as in the single units to establish tuning curves and mutual information between single-voxel responses and viewpoint (respectively identity). All voxels from M6, M7, and M8 were included in this analysis. The results are shown in Figure 2 (bottom) for viewpoint and in Figure 3 (bottom) for identity. Summarizing results for viewpoint tuning, we observe the following. In ML/MF, there is a significant proportion of voxels significantly tuned to each of the five viewpoints (all $p < 0.001$); as in the single units, the tuning curves all present a single peak. The tuning of voxels is well balanced across viewpoints. In AL/AF, we again find a significant proportion of voxels tuned to each of the five viewpoints (all $p < 0.001$, except for R45, $p < 0.05$). Note that the number of voxels in each class is less even, with a predominance of frontal- and full-profile-tuned voxels. The tuning curves are U-shaped, as in the single neurons. In AM, there is a significant proportion of units significantly tuned to L90 and F, with a predominance of units tuned to F; the tuning curves do not significantly depart from chance, but show the expected trends. Therefore, we find some tuning of single voxels to viewpoint, mostly in ML/MF but also in more anterior patches, a pattern broadly consistent with the single-unit results. Turning to single-voxel tuning to identity, we observed a striking dissociation with our electrophysiological data. In ML/MF, there is a significant proportion of voxels significantly tuned to ID3 ($p < 0.01$), ID4 ($p < 0.05$), and ID5 ($p < 0.01$), with a predominance of voxels tuned to ID5. The tuning curves reflect this weakly. In AL/AF and AM, our analyses do not pick up on any significant tuning to identity.

This picture is very different from that which emerged from the analysis of single units—single voxels do not reflect the identity tuning of underlying single units in the anterior face patches (AL/AF and AM), whereas they reflect the viewpoint tuning of the underlying units well in those same face patches. This conclusion is readily apparent from the tuning factors (which measure the

difference between the normalized response to the preferred stimulus and the average normalized response to all other stimuli, in units of SDs), which are depicted in Figures 2 and 3, right. Another interesting observation is that conversely, single voxels in ML/MF are better tuned to identity and viewpoint than single neurons in the same patch.

Decoding invariant representations of viewpoint and identity from single-unit pseudopopulations in the face patches

To further characterize the information at the population level in the single-unit recordings, we attempted to combine information from several neurons linearly. We performed multivariate analyses using a linear Support Vector Machine classifier (LIBSVM implementation, Chang and Lin, 2011). We looked, in turn, for viewpoint information and for identity information in each of the three face patches that were recorded from ML/MF, AL, and AM.

To find evidence for identity-invariant viewpoint information, it was critical to use different identities as training and testing examples (Anzellotti et al., 2014). Therefore, we restricted the training set to four of the five identities and the testing set to the remaining identity; we performed this procedure using each identity as the testing identity in turn, therefore, five training/ testing rounds were run at each cross-validation fold and averaged together (Fig. 4). To assess the significance of the final results we used a permutation test (1000 permutations), which consisted of replicating the entire procedure after randomly shuffling class labels. With 1000 permutations, the highest significance that can arise is $p < 0.001$, corresponding to a situation when no surrogate dataset led to better accuracy than the real dataset. We found that the classifier performs well above chance for viewpoint classification in all three faces patches: ML/MF, $p < 0.001$; AL, $p < 0.001$; AM, $p < 0.001$ (Fig. 5a, left). Deeper insight into the nature of the viewpoint information carried by the pseudopopulations in each face patch can be gained by looking at the pattern of errors made by the linear classifier. Classically, these errors can be represented using confusion matrices; however, confusion matrices only keep track of the final decision made on each testing example without a record of how difficult the decision was. A more complete picture is offered by the average class probability outputs for each class input, which in LIBSVM is computed using an internal fivefold cross validation (Wu et al., 2004) (Fig. 5a, right). Rows in each matrix represent the true labels in the testing set; columns represent the labels that the classifier chooses. ML/MF shows a clear view-specific representation, with some degree of confusion between views that are visually similar, especially the half and full profiles. AL performs slightly better than ML/MF in terms of overall accuracy; although there is less confusion between half and full profile views, the symmetric profile views are hard to tell apart, evincing the emergence of mirror symmetry at this stage of face processing. Finally, AM performs significantly less well than the more posterior patches in disambiguating viewpoint; however, some mirror symmetry remains and the frontal view stands apart from the profile views. Finally, note that, in these analyses, we did not equate the number of cells available from each patch for decoding. Because this can be an issue when comparing performance across loci, we performed the same analyses using only 40 randomly selected neurons in each patch; the results are qualitatively similar (ML/MF accuracy = 51.7%, AL accuracy = 55.7%, AM accuracy = 31.4%; all $p < 0.001$).

We used a similar scheme to quantify viewpoint-invariant identity information in the three face patches, restricting the
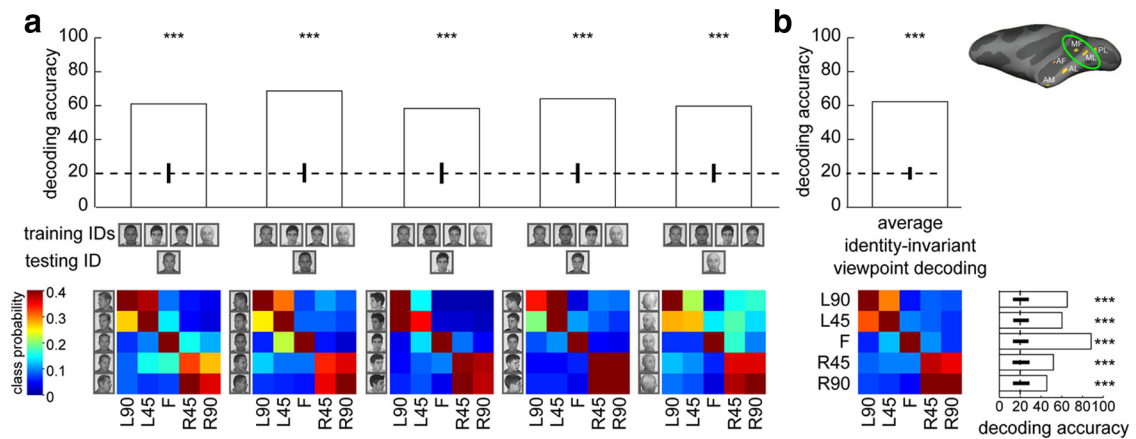
Figure 4. Example of the invariant decoding procedure: decoding viewpoint in the single-unit data (monkeys M1, M2, and M3) from patch ML/MF. ***a***, Each identity is used in turn for testing while a linear SVM is trained on examples from the other four identities; accuracies are shown in the top bar plot and the corresponding probability estimates are shown below. ***b***, Average accuracy and probability matrix. Note that the frontal viewpoint is easier to label than profile views (row-wise accuracy, bottom right). On each bar plot, the thick vertical line centered at chance level (20% correct) is the 95% confidence interval as estimated by a permutation test (1000 repetitions); *p*-values are derived from those null distributions (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).
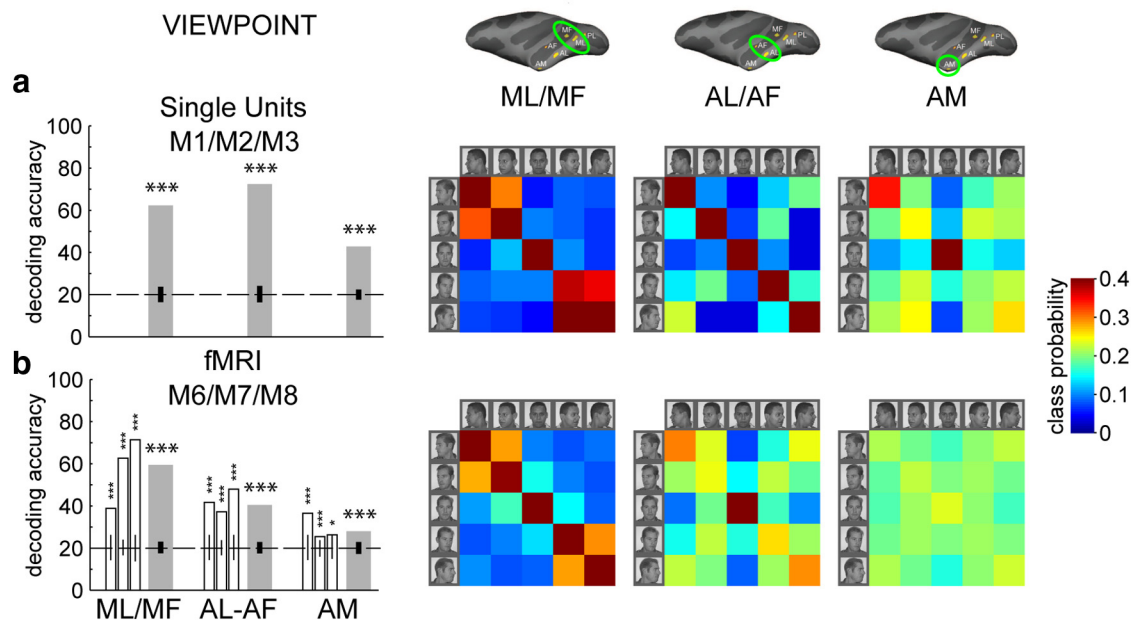

Figure 5. Comparison of identity-invariant viewpoint decoding with single-unit and fMRI data. ***a***, Single-unit data. Left, Decoding accuracy in the three face patches. Chance is indicated with a dashed black line. The 95% interval from a permutation test is shown as a vertical line for each patch. Right: probability estimates for each patch. Rows represent the true labels (L90, L45, F, R45, R90) and columns represent the labels that the classifier chooses from. ***b***, fMRI data. The results are averaged across the three monkeys M6, M7, and M8 (left, shaded bars; right, probability matrices). Individual results are also shown (left, empty bars). Null distributions are estimated by permutation tests (1000 repetitions) (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

training set to four of the five viewpoints and using the fifth viewpoint for testing. We performed this procedure using each viewpoint for testing in turn; therefore, five training/testing rounds were run at each cross-validation fold and averaged together. We found that all three face patches have enough information to discriminate between the five identities significantly above chance level: ML/MF, $p < 0.001$; AL, $p < 0.001$; AM, $p < 0.001$ (Fig. 6a, left), although the accuracy increases greatly from posterior to anterior regions. The output probability matrix (Fig. 6a, right) for ML/MF shows that the above-chance performance is driven mainly by the fifth identity being easily distinguished from the other four. This brought our attention to low-level differences in the image set, which we further investigate in a later section looking at fMRI decoding in the early visual cortices. In

AL, significant decoding is achieved for each identity. It is more difficult for the classifier to generalize to the frontal view from profile views (accuracy for testing viewpoint $F = 34.2\%$) than to generalize to another profile view (accuracy for testing viewpoint L90 $= 57.1\%$, L45 $= 53.3\%$, R45 $= 49.2\%$, R90 $= 60.4\%$), as predicted by mirror symmetric tuning. Performance in AM is very high; a simple linear classifier applied to the population of AM neurons thus achieves viewpoint-invariant face recognition. We found that these results hold when considering the complete set of 25 identities from Freiwald and Tsao (2010) (ML/MF accuracy $= 5.8\%$, AL accuracy $= 21.4\%$, AM accuracy $= 43.0\%$; all $p < 0.001$ except for ML/MF $p = 0.03$). Finally, as noted previously, we selected 40 units at random in each patch and found that the results are qualitatively similar,
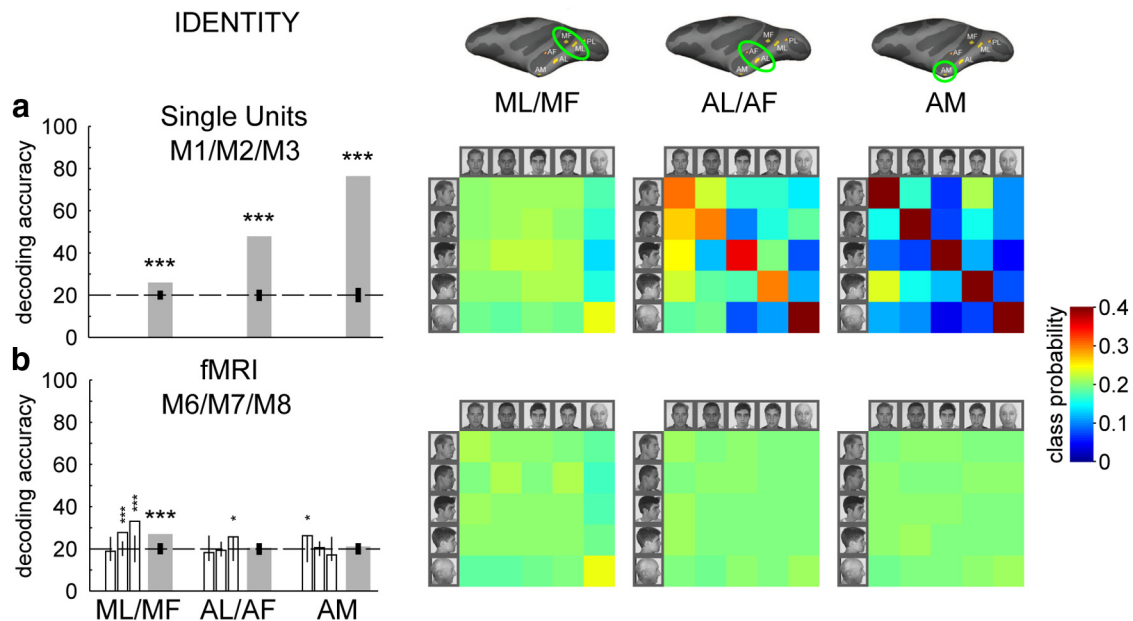
Figure 6. Comparison of viewpoint-invariant identity decoding with single-unit and fMRI data. *a*, Single-unit data. Left, Decoding accuracy in the three face patches. Chance is indicated with a dashed black line. The 95% interval from a permutation test is shown as a vertical line for each patch. Right, Probability estimates for each patch. Rows represent the true labels (ID1, ID2, ID3, ID4, ID5) and columns represent the labels that the classifier chooses from. *b*, fMRI data. The results are averaged across the three monkeys M6, M7, and M8 (left, shaded bars; right, probability matrices). Individual results are also shown (left, empty bars). Null distributions are estimated by permutation tests (1000 repetitions) (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

with a slightly decreased performance in the anterior patches (ML/MF accuracy = 24.9%, AL accuracy = 34.7%, AM accuracy = 48.7%; all $p < 0.001$).

fMRI decoding of viewpoint retrieves the information present in the single-unit populations
In keeping with the analyses performed on the single-unit data, we attempted to classify viewpoint with a linear SVM, training on four of the five identities and testing on the remaining one and repeating this procedure with each identity as the test identity. We implemented a leave-one-run-out cross validation scheme to ensure complete independence between the training and testing examples. We present the results averaged across the three monkeys M6, M7, and M8 in Figure 5b. We found that the linear SVM classifier performs above chance for viewpoint classification in all three face patches, on average across M6, M7, and M8 (all $p < 0.001$). Critically, we found that the probability matrices (Fig. 5b, right) from the fMRI data are in very good agreement with the probability matrices that we get from the single-unit data (Fig. 5a, right). Of particular interest is the emergence of mirror symmetry in AL/AF, as described in the single-unit pseudopopulation data. To quantify the match, we computed a Spearman correlation between the probability matrices derived from the single-unit data and from the fMRI data. We found that the patterns for each face patch are significantly correlated (ML/MF $\rho = 0.76$, AL/AF $\rho = 0.77$, AM $\rho = 0.60$; without diagonal elements, ML/MF $\rho = 0.57$, AL/AF $\rho = 0.59$, AM $\rho = 0.51$, all $p < 0.001$). This strikingly demonstrates the ability of fMRI MVPA to reveal the tuning functions of neurons contained within a region of the cortex.

fMRI decoding of identity fails to retrieve information from the anterior face patches AL/AF and AM
We conducted a similar analysis to the one described for viewpoint decoding. We left out each viewpoint in turn for testing and trained the classifier to decode identity from the four other view-

points. The results are presented as in the previous section, averaged across M6, M7, and M8, in Figure 6b. In the single-unit data, we see that performance improves dramatically from posterior patches to anterior patches, a pattern that we do not find in the fMRI data. In fact, we can only decode identity above chance in ML/MF and, as in the single-unit data, performance is driven up by ID5 (Fig. 6a; ML/MF $p < 0.001$, AL/AF: $p = 0.382$, AM, $p = 0.289$). A Spearman correlation analysis between the probability matrices likewise fails to find similar patterns between the single-unit and fMRI decoding (ML/MF $\rho = 0.17$, AL/AF $\rho = 0.14$, AM $\rho = 0.13$, all $p > 0.05$; without diagonal elements, ML/MF $\rho = 0.16$, AL/AF $\rho = 0.05$, AM $\rho = 0.17$, all $p > 0.05$).

Anterior face patches have lower functional SNRs
fMRI typically yields noisier measurements in the anterior temporal lobes than in more posterior cortical areas; because invariant identity information lies mostly in anterior areas whereas invariant viewpoint information lies in posterior areas, this could partly explain the discrepancy. We quantified the functional SNR (fSNR, which we defined within a GLM framework as the average magnitude of the fMRI signal change divided by the SD of the residuals across time) for each patch in all five monkeys. We found that fSNR significantly differs between face patches for each monkey (one-way ANOVAs; all $p < 0.05$). Multiple-comparison tests (using Tukey's honestly significant difference criterion) showed that fSNR is significantly lower in AM than in AL/AF (all monkeys) and in AM than in ML/MF (all but M6). fSNR in AL/AF is either significantly lower (three of five monkeys), not statistically different (M6) or significantly higher (M4) compared with fSNR in ML/MF. It is thus possible that a relatively low fSNR hindered our ability to read out identity information in the anterior face patches; however, because we did find significant decoding of viewpoint information in AL/AF (and in AM), the lack of identity decoding in AL/AF (and in AM) is not solely due to a low SNR in that area. Understanding the failure of fMRI in retrieving identity information in anterior face patches thus requires a
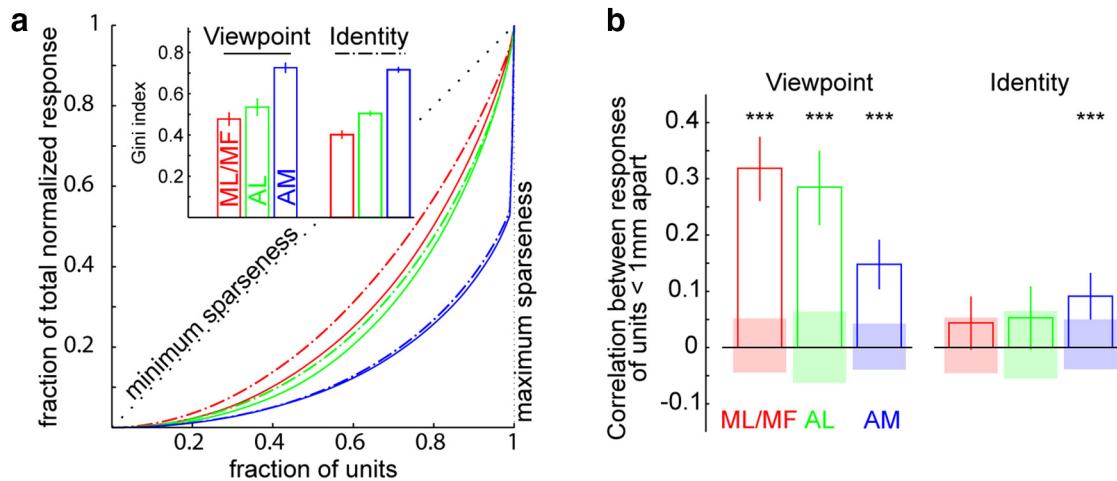
Figure 7. Properties of the single-unit population coding of viewpoint and identity. ***a***, Sparseness of the neuronal representations of viewpoint (solid lines) and identity (dash-dotted lines) increases from ML/MF to AM. The Gini index (bar plot) corresponds to twice the area below the diagonal when plotting the fraction of the total response against the fraction of units. ***b***, Clustering of single-unit selectivity to viewpoint and identity. The correlation of responses of neighboring units (≤1 mm) was assessed as a proxy for clustering, across all viewpoints and across all identities in the face views image set (Freiwald and Tsao, 2010) in the three regions of interest. Viewpoint selectivity in ML/MF and AL is highly clustered, whereas identity selectivity does not show above chance clustering. In AM, both viewpoint and identity selectivity are clustered, but to a much lesser extent than in ML/MF or AL. A 95% confidence interval for the distribution of chance was estimated with a permutation test (shaded areas). Error bars are SEM (across all pairs of units, i.e., 385 in ML/MF, 326 in AL/AF and 610 in AM).

more in depth investigation of the properties of the neural population representations of identity and viewpoint.

Neural population representations of viewpoint and identity: sparseness and clustering

The signal measured in fMRI is hemodynamic and a prerequisite for a sizeable hemodynamic response is that enough neurons be active in a given area in response to a stimulus. Sparseness reflects the proportion of a neural population that is active in response to a stimulus. We used the Gini index (Hurley and Rickard, 2009) as a measure of sparseness: if only one neuron in a population responds to a given stimulus, the Gini index is 1, whereas if all neurons respond at the same level (compared with their maximal response), the Gini index is 0. We found that the sparseness of both identity and viewpoint representations (obtained by averaging single image responses across viewpoints and identities, respectively, before computing the Gini index) increases significantly from ML-MF to AM (one-way ANOVA: viewpoint, $F_{(2,21)} = 199.4$, $p = 2 \times 10^{-14}$; identity, $F_{(2,72)} = 3477.39$, $p = 2 \times 10^{-72}$; Fig. 7a); therefore, neuronal responses in anterior face patches are sparser than in posterior face patches. The critical comparison to make is between the representations of viewpoint and identity in AL, the face patch where identity and viewpoint information are both very well represented in the single units but only viewpoint can be retrieved with fMRI MVPA. There, we found that viewpoint representations are in fact slightly sparser than identity representations (two-sided t test, $T(31) = 3.36$, $p = 0.002$), ruling out sparseness as the main factor preventing identity decoding in AL.

If units with similar tuning are scattered rather than concentrated, it is also less likely that the hemodynamics will carry information about the underlying representations. Retrieving the exact location of the recording electrode and comparing it across different sessions is not possible in our setup (fine electropolished Tungsten electrodes were inserted anew, through a grid hole, at each recording session; for details, see Freiwald and Tsao, 2010). However, we can be confident in comparing locations sampled during the same session (one electrode penetration, at different depths). Although this does not allow us to recover the full to-

pography of viewpoint and identity selectivities, we can compare the response profiles of nearby units (within 1 mm) recorded on the same day (Fig. 7b). Clustering for viewpoint is significantly higher than chance in all face patches (all $p < 0.001$); clustering for identity is also higher than chance in AM ($p < 0.001$) and shows a strong trend for ML/MF ($p = 0.055$) and AL $p = 0.059$. A two-way ANOVA, with stimulus dimension and face patch as factors, showed a main effect of stimulus dimension ($F_{(1,2636)} = 66.93$, $p = 0$), indicating that clustering for viewpoint is higher than for identity. There was also a significant interaction ($F_{(2,2636)} = 8.56$, $p = 0.0002$) corresponding to opposite trends of descending and ascending clustering from posterior to anterior patches, depending on whether viewpoint or identity was considered. Critically, clustering in AL for viewpoint is much higher than clustering of identity tuning (planned t test, $T(650) = 4.8632$, $p = 10^{-6}$). This is likely to play a major role in the discrepancy between viewpoint and identity decoding in AL with fMRI data. Note also that, in AM, clustering is weak for both viewpoint and identity, which may account for why fMRI data rather poorly reflects the information available in the single-unit pseudopopulations in AM.

Face viewpoint and identity information outside of the face patches

Our stimuli were not perfectly equated in terms of low-level properties (see Discussion). One way to assess the extent of the low-level confounds is to look at how well we can decode identity-invariant viewpoint and viewpoint-invariant identity in early visual areas. We mapped the early visual areas for M6, M7, and M8 using a meridian mapping paradigm (Sereno et al., 1995; Fize et al., 2003). The borders of V1, V2, V3, and V4 were hand-drawn on the computationally flattened occipital cortices along the highest absolute values for the contrast "vertical − horizontal"; the numbers of voxels for each ROI are reported in Table 1. We performed decoding of face identity and viewpoint in each of these early visual areas using the same procedures described previously for decoding in the face patches. On average, across M6, M7, and M8, we found that we can decode both viewpoint and identity significantly above chance in all early visual areas (Fig. 8;
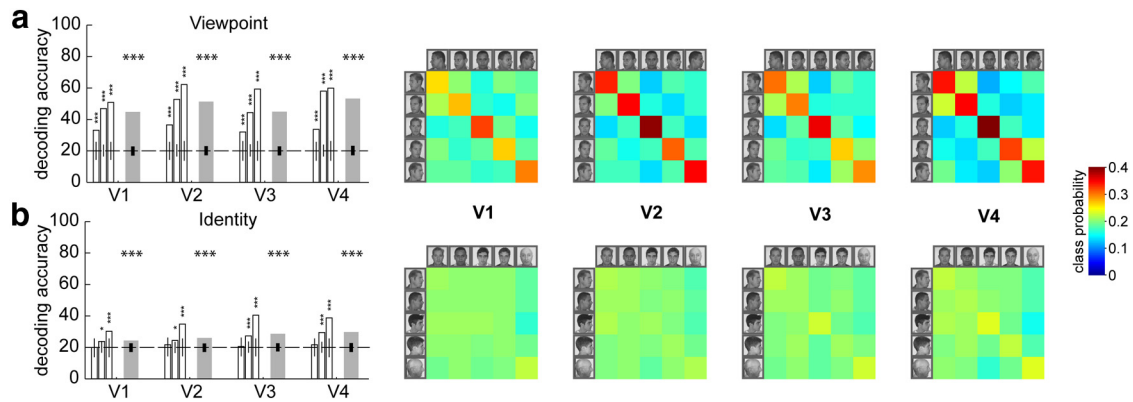
Figure 8. fMRI decoding results in the early visual areas (V1, V2, V3, and V4). ***a***, Viewpoint decoding. The results are averaged across the three monkeys M6, M7, and M8 (left, shaded bars; right, probability matrices). Individual results are also shown (left, empty bars). Null distributions are estimated by permutation tests (1000 repetitions) (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). ***b***, Identity decoding.

all $p < 0.001$), attesting to the presence of low-level cues. For viewpoint decoding, this is not unexpected. It is, however, more surprising for identity decoding; the low-level confounds appear to be best captured at the level of V3 and V4 (note that the decoding of identity in ML/MF is worse than in V4; Fig. 5*b*).

Finally, it is natural to wonder whether fMRI MVPA can retrieve viewpoint and identity information beyond early visual areas and outside of the face patches in our experiments. One of the strengths of fMRI as a brain imaging technique is that the whole brain is recorded from at each time point. We thus ran a searchlight decoding (information mapping, Kriegeskorte et al., 2006) procedure to probe other parts of the brain for (identity-invariant) viewpoint and (viewpoint-invariant) identity information. Information maps were thresholded at a binomial p-value of 0.0001 (uncorrected) for visualization. We found that viewpoint information is present in much of the posterior brain, including early visual areas (the results for M7 are shown in Fig. 9, top). Identity information is retrieved above chance in posterior areas in some monkeys, but not in anterior areas where the invariant representation of identity is expected from the single units (results for M7 are shown in Fig. 9, bottom). The searchlight analysis also serves as a control for the effect that the number of voxels available to the decoding algorithm may have in accounting for differences across face patches. Because there are more voxels in ML/MF than in AL/AF or AM (Table 1), it could be argued that chance performance in AM for decoding identity is due to the relatively low number of voxels. Because the searchlight analysis uses the same number of voxels throughout the brain (here, 125) and because we find significant decoding of identity around M7's ML/MF, but not around AM, this shows that the number of voxels is not the limiting factor.
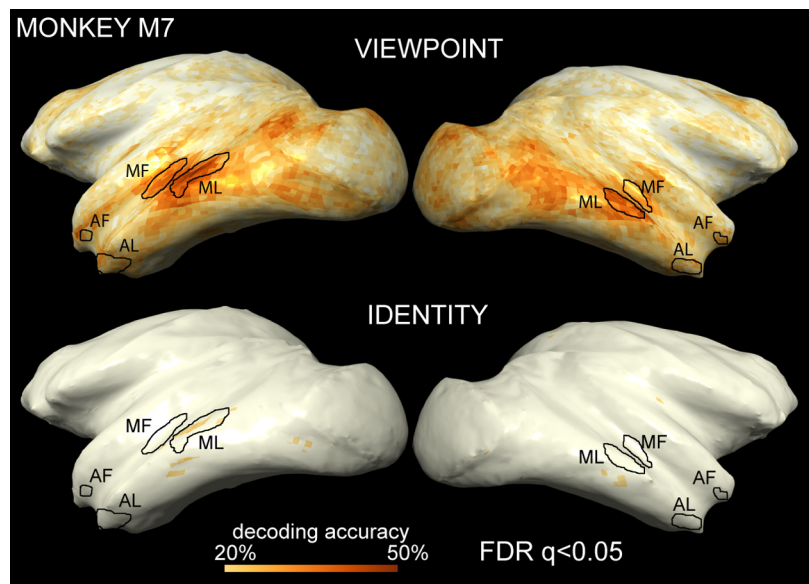

Figure 9. Searchlight decoding of viewpoint and identity for monkey M7 projected on the lateral cortical surface (medial and ventral surface not shown, so AM cannot be seen). ***a***, Viewpoint decoding. A cubic searchlight of dimensions $5 \times 5 \times 5$ voxels is centered at each voxel in turn in the whole brain and a decoding procedure identical to the one described for the face patches is applied. The decoding accuracy for each searchlight is mapped to the center voxel. The decoding map is thresholded at a binomial p-value of 0.0001 (uncorrected) for visualization. The face patches (from an independent functional localizer, see text) are outlined in black to evidence the overlap with searchlight decoding accuracy. There are many areas of the brain from which viewpoint information is retrieved with a high accuracy. ***b***, Identity decoding. Decoding accuracy is poor; only a few clusters are visible at this threshold (including one in the left ML); no significant voxels on the ventral/medial surface. Projections were generated with pycortex v.0.1.1 (https://github.com/gallantlab/pycortex).

Discussion

fMRI decoding has been applied to a wide array of research questions, from basic vision to decision making (Tong and Pratte, 2012). It is believed to provide a tool for fMRI researchers to gain insight into the fine-scale informational content of brain regions (Kriegeskorte and Bandettini, 2007). Here, we tested the ability of fMRI MVPA to reveal information encoded in the underlying populations of neurons in the macaque monkey. This is the first comparison of fMRI MVPA results to underlying population representations in the same species using the same stimuli.

Single-unit recordings targeted with fMRI in the macaque have shown that units in the different face patches are differentially tuned to viewpoint and identity (Freiwald and Tsao, 2010).

We applied multivariate techniques to gain further insight into the informational content of each patch. Then, using a high-quality fMRI dataset, we demonstrated a discrepancy between the successful retrieval of viewpoint information with MVPA as predicted from the single-unit data and the failed retrieval of identity information. Weak clustering of like-tuned units may have contributed to the failure of identity information retrieval.

Our analysis of clustering was limited by the precision with which we could localize the recording sites. Critically, the fine electrodes used to record from single units were removed after each recording session, preventing a precise estimate of the distance between neurons recorded from in different sessions. We thus only considered units recorded from in the same session, with the same electrode positioned at different depths throughout the session. Penetration angles were determined to avoid overlying blood vessels while still hitting the target face patch; they were not designed to be exactly orthogonal to the skull or to the cortical surface. Although we obtained MRI of the electrode for every penetration and could compute the angle with respect to the cortical surface, our clustering argument is based on distance between recorded units with no specification of whether these units belong to the same or to different cortical columns. It would be a worthy next step to use methods such as those described in Issa et al. (2013) to recover the full topography of cells with different selectivities to identity and viewpoint in the face patches.

In the single-unit experiment, a large number of faces (200) were shown in rapid succession (200 ms each, 200 ms blank), whereas only 25 faces were shown for 24 s each in the fMRI experiments (with 24 s blanks). Boredom, attentional differences, and neural adaptation can be raised as factors hindering our ability to decode identity information in the fMRI experiments. The timings of the fMRI experiment cannot be matched to the single-unit experiment due to the timescale of the hemodynamic response. Instead, as a control, we recorded from single units while matching the presentation times of the fMRI experiments. We collected data from six AM cells while M5 viewed 24-s-long presentations of the five identities in our dataset (frontal views), with spatial jittering as in the fMRI paradigm. We averaged the ranked responses (based on the response elicited by each stimulus at short latency, 0–1000 ms) of all six units to the five stimuli: although the responses showed adaptation (the firing rate to the preferred ID in the 23–24 s period was about 65% of the initial rate), it was also evident that identity tuning remained throughout the trial (paired t test between the firing rate to the best ID and the firing rate to the second best ID in the 23–24 s window: T(5) = 6.30, $p = 0.001$). Other trivial differences between the single-unit and the fMRI paradigm (e.g., stimulus size, spatial jitter) do not affect the conclusions we draw here, given the established size and position invariance of facial coding in anterior brain patches (e.g., supplemental Fig. 10 in Freiwald and Tsao, 2010).

The fMRI and the single-unit experiments discussed here were conducted in different animals. We know that the locations and properties of the face patch system are very reproducible from one monkey to the next (Freiwald and Tsao, 2010), warranting between-animal comparisons. As a further control, we collected single-unit data from M5 's AM (53 units) using the same image set. We confirmed, using the same linear classification procedure, that AM units in M5 carried viewpoint-invariant identity information (accuracy = 27.6%, $p < 0.001$) and identity-invariant viewpoint information (accuracy = 58.1%, $p < 0.001$).

The information derived from the single-unit pseudopopulations with a linear classifier is an impoverished depiction of the information truly present in the face patches. A faithful account of the information encoded in these populations would require a completely different setup, such as simultaneous recordings from many units in the population to exploit trial-to-trial covariance. In addition, the firing rate of each unit is not a sensitive measure of what information the brain is processing—critical information is represented in the precise timing of spikes as well as in the subthreshold postsynaptic potentials. Accordingly, we do not claim that the results reported here represent the best that can be done with single units, nor that we have definitively characterized the information present in each face patch: our results are a lower-bound estimate of the information that is present. The crux of our approach is to then look at the fMRI data to determine whether that (impoverished) information can be recovered.

Equating low-level differences is often an important step in vision science experiments. The image set used here consisted of grayscale pictures that had not been further processed. For example, images of ID5 appearing brighter than those of the other individuals. We argue here that these low-level differences (which we picked up on in early visual areas) do not affect the main conclusions. First, we find that the different face patches represent differential information about the faces; if low-level differences were the only source of information, one would expect to find a similar pattern of decoding in all regions of interest. Second, we find that no identity information can be read out from AL/AF or AM in the fMRI data; the low-level differences should have helped us to detect identity information, but even in these conditions, we were unable to. Finally, we are comparing decoding accuracies between fMRI and single-units and, if low-level cues accounted entirely for both, then we should have found decoding accuracies obtained using the two techniques to be the same.

There have been several reports of significant fMRI decoding of face identity in the human literature (Kriegeskorte et al., 2007; Natu et al., 2010; Nestor et al., 2011; Anzellotti et al., 2014). Our results in this study may appear at odds with these positive reports. However, there are many factors that differentiate our study from these. First, our experiments were conducted in a different species. We show that there is enough information in the single units in anterior faces patches to retrieve the identity of the human faces that are presented to the monkeys; however, the properties of the neural populations that encode face identity may differ between macaques and humans, accounting for why fMRI MVPA may fail in one case and succeed in the other. Another difference is that in our experiments the monkeys' task was passive fixation, whereas there is invariably a task enforcing attention to face identity in the human experiments; whereas the single-unit recordings demonstrate that neurons at the top of the visual hierarchy (AM neurons) encode identity in an invariant manner under passive viewing conditions, a task enforcing attention to the images may be critical for this representation to be sustained and yield a significant fMRI signal over the course of a long block. Another critical difference is that semantic information is generally associated with the faces in human fMRI experiments either imposed by the experimenter in the design (e.g., each face has a name, a job) or self-generated by the human subject to facilitate their task (e.g., "this is the guy that looks like my friend Joe"). The representation of identity thus becomes much richer than that resulting from bottom-up face processing, which may help to generate discriminable fMRI patterns. It is unknown whether this happens for the macaques.

In sum, our study validated the notion that fMRI MVPA is a powerful tool for fMRI analysis. fMRI MVPA retrieved information about facial viewpoint with high fidelity, for example, extracting a mirror symmetric representation in AL/AF (Kietzmann et al., 2012; see also Axelrod and Yovel, 2012). However, we also unveiled a key limitation of fMRI MVPA in its failure to retrieve information about face identity in regions where single-unit recordings demonstrated that this information was represented in the underlying neuronal populations. Further studies are needed to elucidate the relationship between information decodable from fMRI multivoxel patterns versus single-unit populations for other variables in other brain regions. Our results underscore the point that the success of fMRI decoding depends strongly on the particular spatial organization of the variable being decoded. We suspect that there are many other variables such as facial identity that do not show a strong spatial topography and are unlikely ever to be successfully decoded by fMRI multivoxel pattern analysis.

Notes

Supplemental material for this article is available at http://tsaolab. caltech.edu/?q=supp_material. Results of control experiments and analyses; individual fMRI results for M4-M8. This material has not been peer reviewed.

References

Anzellotti S, Fairhall SL, Caramazza A (2013) Decoding representations of face identity that are tolerant to rotation. Cereb Cortex 24:1988–1995. CrossRef Medline

Axelrod V, Yovel G (2012) Hierarchical processing of face viewpoint in human visual cortex. J Neurosci 32:2442–2452. CrossRef Medline

Chang CC, Lin CJ (2011) LIBSUM: a library for support vector machines. ACM Transact Intel Syst Technol 2:1–27.

Cusack R, Brett M, Osswald K (2003) An evaluation of the use of magnetic field maps to undistort echo-planar images. Neuroimage 18:127–142. CrossRef Medline

Fize D, Vanduffel W, Nelissen K, Denys K, Chef d'Hotel C, Faugeras O, Orban GA (2003) The retinotopic organization of primate dorsal V4 and surrounding areas: a functional magnetic resonance imaging study in awake monkeys. J Neurosci 23:7395–7406. Medline

Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. Science 330:845–851. CrossRef Medline

Gratton C, Sreenivasan KK, Silver MA, D'Esposito M (2013) Attention selectively modifies the representation of individual faces in the human brain. J Neurosci 33:6979–6989. CrossRef Medline

Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686–691. CrossRef Medline

Hurley N, Rickard S (2009) Comparing measures of sparsity. IEEE Trans Inf Theory 55:4723–4741. CrossRef

Issa EB, Papanastassiou AM, DiCarlo JJ (2013) Large-scale, high-resolution neurophysiological maps underlying fMRI of macaque temporal lobe. J Neurosci 33:15207–15219. CrossRef Medline

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685. CrossRef Medline

Kietzmann TC, Swisher JD, König P, Tong F (2012) Prevalence of selectivity for mirror-symmetric views of faces in the ventral and dorsal visual pathways. J Neurosci 32:11763–11772. CrossRef Medline

Kriegeskorte N, Bandettini P (2007) Analyzing for information, not activa-

tion, to exploit high-resolution fMRI. Neuroimage 38:649–662. CrossRef Medline

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868. 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

Ku SP, Tolias AS, Logothetis NK, Goense J (2011) fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. Neuron 70:352–362. 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

Natu VS, Jiang F, Narvekar A, Keshvari S, Blanz V, O'Toole AJ (2009) Dissociable neural patterns of facial identity across changes in viewpoint. J Cogn Neurosci 22:1570–1582. 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

Noirhomme Q, Lesenfants D, Gomez F, Soddu A, Schrouff J, Garraux G, Luxen A, Phillips C, Laureys S (2014) Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions. Neuroimage Clin 4:687–694. CrossRef Medline

Ohayon S, Tsao DY (2012) MR-guided stereotactic navigation. J Neurosci Methods 204:389–397. CrossRef Medline

Pereira F, Botvinick M (2011) Information mapping with pattern classifiers: a comparative study. Neuroimage 56:476–496. 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. Medline

Rajimehr R, Bilenko NY, Vanduffel W, Tootell RB (2014) Retinotopy versus face selectivity in macaque visual cortex. J Cogn Neurosci 26:2691–2700. CrossRef Medline

Schreiber K, Krekelberg B (2013) The statistical analysis of multi-voxel patterns in functional imaging. PLoS One 8:e69328. CrossRef Medline

Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. Neuroimage 44:223–231. CrossRef Medline

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893. CrossRef Medline

Tong F, Pratte MS (2012) Decoding patterns of human brain activity. Annu Rev Psychol 63:483–509. CrossRef Medline

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. Nat Neurosci 6:989–995. CrossRef Medline

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670–674. 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

Vanduffel W, Zhu Q, Orban GA (2014) Monkey cortex through fMRI glasses. Neuron 83:533–550. CrossRef Medline

Wu T, Lin C, Weng R (2004) Probability estimates for multi-class classification by pairwise coupling. Journal of Machine Learning Research 5:975–1005.

Zeng H, Constable RT (2002) Image distortion correction in EPI: comparison of field mapping with point spread function mapping. Magn Reson Med 48:137–146. CrossRef Medline