Systems/Circuits

Intelligent Information Loss: The Coding of Facial Identity, Head Pose, and Non-Face Information in the Macaque Face Patch System

Ethan M. Meyers,[1,2] Mia Borzello,[2,3] Winrich A. Freiwald,[2,4] and Doris Tsao[5]

[1]School of Cognitive Sciences, Hampshire College, Amherst, Massachusetts 01002, [2]The Center for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, [3]Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, [4]Laboratory of Neural Systems, The Rockefeller University, New York, New York 10065, and [5]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91126

Faces are a behaviorally important class of visual stimuli for primates. Recent work in macaque monkeys has identified six discrete face areas where most neurons have higher firing rates to images of faces compared with other objects (Tsao et al., 2006). While neurons in these areas appear to have different tuning (Freiwald and Tsao, 2010; Issa and DiCarlo, 2012), exactly what types of information and, consequently, which visual behaviors neural populations within each face area can support, is unknown. Here we use population decoding to better characterize three of these face patches (ML/MF, AL, and AM). We show that neural activity in all patches contains information that discriminates between the broad categories of face and nonface objects, individual faces, and nonface stimuli. Information is present in both high and lower firing rate regimes. However, there were significant differences between the patches, with the most anterior patch showing relatively weaker representation of nonface stimuli. Additionally, we find that pose-invariant face identity information increases as one moves to more anterior patches, while information about the orientation of the head decreases. Finally, we show that all the information we can extract from the population is present in patterns of activity across neurons, and there is relatively little information in the total activity of the population. These findings give new insight into the representations constructed by the face patch system and how they are successfully transformed.

Key words: face processing; face recognition; macaque; neural decoding; population coding

Introduction

Primates are uniquely adept at extracting information from faces (Bruce and Young, 1986; Pokorny and de Waal, 2009). Electrophysiological studies in macaque monkeys have shown that images of faces selectively drive a subset of neurons in the inferotemporal cortex (Perrett et al., 1982; Desimone et al., 1984; Kobatake and Tanaka, 1994), and PET and fMRI studies of humans have shown that there are localized regions that respond more to images of faces than to images in other object classes (Sergent et al., 1992; Kanwisher et al., 1997). More recent work has connected these findings by showing that macaque inferotemporal cortex contains six discrete face-selective areas ("face patches") that have stronger fMRI responses to faces than to nonface objects, and that neurons in these patches exhibit much higher firing rates to images of faces (Tsao, 2006; Freiwald and Tsao, 2010; Issa et al., 2013). Additionally, these face patches form an interconnected network, and neurons in different patches have different response properties (Moeller et al., 2008; Freiwald and Tsao, 2010; Issa and DiCarlo, 2012). Since the face patches form a connected system, they provide an ideal opportunity to systematically and quantitatively compare the amount of information in each patch about different classes of visual stimuli to understand how population codes are hierarchically transformed in the brain.

In this study we used neural population decoding to better quantify the information content in three of these face patches [the middle lateral and middle fundus patches (ML/MF), the anterior lateral patch (AL), and the anterior medial patch (AM)]. In particular, we were interested in two questions. First, do face patches contain information exclusively about faces or do they also contain information about nonface objects? While it is known that neurons inside face patches exhibit higher firing rates to faces than to nonface objects (Tsao, 2006; Issa et al., 2013), in

A   Face-object-body (FOB) stimulus set

B   Image similarity within each category on the FOB stimulus set

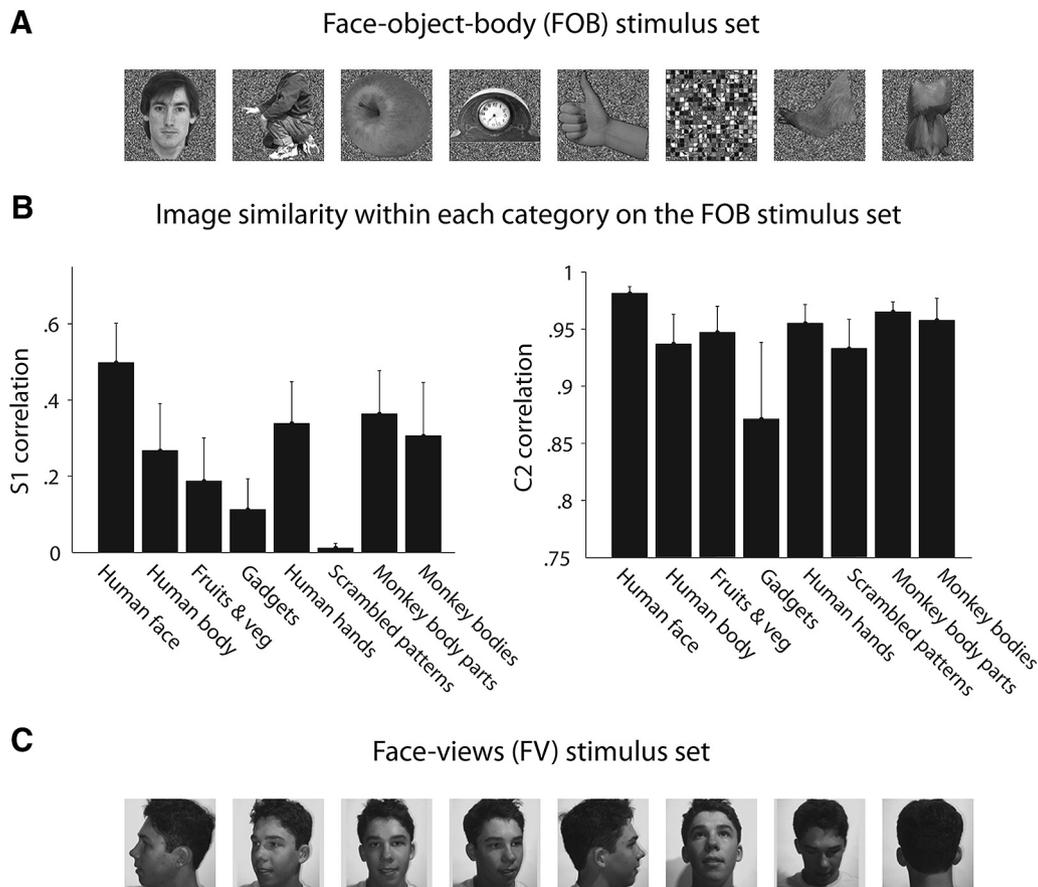**C**   Face-views (FV) stimulus set

Figure 1. Stimulus sets used in the experiments. ***A***, One example of each image from the eight categories in the FOB stimulus set. Each category contained 16 images, for a total of 128 images in this stimulus set. ***B***, Average HMAX S1 and C2 image similarity (Pearson's correlation) between all images in each of the eight categories in the FOB dataset, averaged over all pairs of the 16 images in each category. HMAX S1 features are responses to Gabor filters and thus mimic properties of V1, while C2 features are modeled on higher level visual responses (Serre et al., 2007; Mutch et al., 2010). Error bars show the SD in the HMAX feature correlations across pairs of the 16 images. As can be seen, images in the face category were most similar to each other while images in the scrambles image category were most different from each other. ***C***, An example of the eight head orientations for 1 of the 25 people in the FV stimulus set.

general it has not yet been demonstrated whether neurons primarily code information for distinguishing different images only in high firing rate responses (i.e., "preferred stimuli") or whether information is carried in smaller modulations in the lower range of spiking activity. Thus, it is possible that neurons in these patches contain nonface information in lower firing rate regimes.

Second, we were interested in the extent to which face patches contain information about differences within the domain of faces. Specifically, we explored whether neural populations within the three face patches contain pose-invariant identity about specific individuals, and whether they contain information about the pose of the head. Previous analyses examining single neuron responses have suggested that the more anterior patches are more pose invariant and the more posterior patches contain more information about head pose (Freiwald and Tsao, 2010). By using decoding analyses, however, we are able to better compare pose-specific and pose-invariant identity information carried by the whole-cell population. These analyses give insight into how two important facial variables, pose and identity, are represented at the population level along the face patch system.

Materials and Methods

All procedures conformed to local and U.S. National Institutes of Health guidelines, including the U.S. National Institutes of Health Guide for Care and Use of Laboratory Animals, regulations for the welfare of ex-

perimental animals issued by the German federal government, and stipulations of local authorities.

Stimuli and experimental design

Two stimulus sets were used in the experiments (Fig. 1; Freiwald and Tsao, 2010). The first stimulus set (Fig. 1A), called the "face-object-body" stimulus set (FOB), consisted of 16 images from eight different categories (human faces, human bodies, fruits and vegetables, gadgets, human hands, scrambled patterns, monkey body parts, and monkey whole bodies) for a total of 128 images. To understand how low level image features could affect our results on the FOB dataset, we calculated the HMAX S1 and C2 image similarity of the eight categories by computing Pearson's correlation coefficient between the S1 (or C2) values for all pairs of the 16 images in each category. S1 features are based on Gabor functions at different scales and locations, and are supposed to model V1 simple cell responses, while C2 features are modeled on higher level visual responses (Serre et al., 2007; Mutch et al., 2010). Figure 1B displays the correlation coefficient averaged over all pairs of images in each category and the error bars are the SDs over all pairs within each category.

The second stimulus set (Fig. 1C), called the "face-views" stimulus set (FV), consisted of human face images taken from eight different angles (left full profile, left three-quarter profile, straight, right three-quarter profile, right full profile, up, down, and back) from 25 different individuals. The 25 individuals in the image set spanned a wide range of visual appearances (i.e., different races, genders, ages, hair styles, etc.) and were each presented against a gray background. The monkeys had experience

interacting with 4 of the 25 individuals in this stimulus set, and had not interacted with the remaining 21 individuals.

For both stimulus sets, images were presented to the monkeys in random order in a rapid sequence where each image was shown for 200 ms followed by a 200 ms blank interval. All stimuli were 7×7 degrees of visual angle in size and presented at the center of the screen. Each image was repeated from 3 to 10 times. Monkeys were given a juice reward for maintaining fixation within a 2.5 degree window at the center for the screen for 3 s.

Neural recordings

Single-unit recordings were made from three male rhesus macaque monkeys (*Macaca mulatta*). Before recording, face patches were localized by conducting fMRI scans from each alert monkey and identifying the regions that responded significantly more to images of faces compared with images of bodies, fruits, objects, hands, and scrambled patterns. Neural recordings were performed in four of these face patches (two middle patches, ML and MF, and two anterior patches, AL and AM). Following the practice of Freiwald and Tsao (2010) who found similar response properties in ML and MF, we combined the data from these patches. Thus our analyses examine three face-selective regions—ML/MF, AL, and AM—which appear to be organized hierarchically, with ML/MF at the input level, AM at the output level, and AL representing an intermediate level of processing (Freiwald and Tsao, 2010).

Only well isolated single units were used in our analyses. The data used in these analyses was previously presented in Freiwald and Tsao (2010), and a more detailed description of the surgical, recording, and experimental procedures can be found there and in Moeller et al. (2008).

Data analyses

Population decoding methods were used to analyze the data in this paper (for single neuron selectivity index measures see Tsao et al. (2006), Freiwald and Tsao (2010)). In brief, a pattern classifier was trained on the firing rate activity of a population of neurons to discriminate between stimuli. The "decoding accuracy" is then calculated as the percentage of stimuli that are correctly classified using a "test set" of data that consists of the firing rate activity from a different set of trials. All analyses were done in MATLAB using a beta version of the Neural Decoding Toolbox (Meyers, 2013). Below we first describe common parameters that are used for all analyses, and then we describe in more detail the specific parameters of each particular analysis. For more information about these decoding analyses see Meyers and Kreiman (2012) and Meyers (2013).

Parameters common to all analyses. For all analyses, pseudo-populations were created that contained 40 neurons that were randomly sampled from the larger populations of ML/MF, AL, or AM neurons (only neurons that had recordings from at least three repetitions of each stimulus were included in the analyses). Three cross-validation splits of the data were used by randomly selecting three trials for each stimulus for each neuron. Thus for most analyses, the classifier was trained on two R^{40} dimensional vectors from each class and tested on a third vector. The decoding procedure was run 50 times creating different pseudo-populations and cross-validation splits on each run (i.e., 50 resample runs were used). A Poisson Naive Bayes classifier (PNB) was used (see below, Classifiers). All results that are plotted as a function of time (Figs. 3, 4 and 6–8) are based on firing rates in 100 ms bins that were sampled at 10 ms intervals, with the data starting at the time when the stimulus was shown. These 100 ms bins are plotted such that the decoding accuracy is aligned to the center of each bin. For Figures 2 and 5, the average firing rate in a 300 ms bin that started 100 ms after stimulus onset was used. The decoding accuracies reported in the body of the paper are also based on using a 300 ms bin that starts 100 ms after stimulus onset. Since decoding accuracy is generally higher when larger bin sizes are used, the decoding results in this 300 ms bin are higher than the decoding accuracies shown on the 100 ms sliding bin figures (we use the smaller bin sizes when plotting the results as a function of time to better see the temporal evolution of the decoding accuracies). It should be noted that decoding accuracy gives a lower bound on the amount of information in a brain region (Quian Quiroga and Panzeri, 2009); thus, we say there is "information" about a particular variable when the decoding accuracy for that variable is above chance.

Table 1. A table listing the classification decision functions for the four different classifiers used in this paper

A. Poisson Naive Bayes (PNB)
$$\arg \max_{c} \log (\mathbf{w}_c)^T \boldsymbol{x} - n\bar{w}_c$$

B. Total Activity (TAct)
$$\arg \min_{c} |\bar{w}_c - \bar{x}|$$

C. Maximum Correlation Coefficient (MCC)
$$\arg \max_{c} \frac{(\mathbf{w}_c - \bar{w}_c)^T (\boldsymbol{x} - \bar{x})}{\|\mathbf{w}_c - \bar{w}_c\| \|\boldsymbol{x} - \bar{x}\|}$$

D. Minimum Angle (Min Ang)
$$\arg \max_{c} \frac{\mathbf{w}_c^T x}{\|\mathbf{w}_c\| \|\boldsymbol{x}\|}$$

$\mathbf{w}_c$ is a vector that is the mean of the training data from class c, $\bar{w}_c$ is a scalar that is the mean of $\mathbf{w}_c$, $\boldsymbol{x}$ is the test vector to be classified, $\bar{x}$ is a scalar that is the mean of $\boldsymbol{x}$, and n is the number of neurons; thus training the classifier consists of learning $\mathbf{w}_c$ and $\bar{w}_c$ and testing the classifier consists of determining which class $\boldsymbol{x}$ belongs to. As can be seen, all these classifiers are rather similar and mainly differ in how they normalize the data and, consequently, whether they take the overall level of population activity into account (A and B) or whether they only examine relative differences in the firing rate activity between neurons (C and D). TAct, Total Activity; MCC, Maximum Correlation Coefficient; Min Ang, Minimum Angle.

Classifiers. A PNB classifier (Duda et al., 2001) was used for all analyses in this paper. In this classifier, each neuron's spiking activity is modeled as a Poisson random variable, and each neuron's activity is assumed to be independent given a particular stimulus. To explore how the overall activity level of the whole population affected decoding performance (Fig. 8), we used several additional classifiers. The aim of these analyses was not to do an extensive search for the best possible classifier, but rather to examine how the overall firing rate level (which is much higher when faces are shown) affects the decoding accuracy of simple biologically plausible classification rules that could easily be implemented in neural circuits. Such analyses can give insight into whether information is encoded in overall spiking activity level of a population or only in the relative activity patterns across neurons.

To summarize how these classifiers operate, let w_c be a vector that consists of the mean of the training vectors from class c, and let x be a test point to be classified; (thus w_c and x are vectors in R^n where n is the number of neurons, which for all analyses was $n = 40$). Also, let $\bar{w}_c$ and $\bar{x}$ be scalars that are found by taking the mean of the values in the vectors w_c and x, respectively. Then training for the Total Activity classifier consisted of computing the $\bar{w}_c$ for each class, and training the other classifiers consisted of calculating w_c for each class. The decision functions that each classifier uses to classify a test point x are given in Table 1. As can be seen, despite the different interpretations that are often given to these classifiers (e.g., measures of correlations, angles in a high dimensional space, and maximum likelihood estimates), they are all rather similar and mainly differ in the way that they normalize the data. For the Maximum Correlation Coefficient and Minimum Angle classifier, the data was z-score normalized using the mean and SDs of the training data to prevent neurons with higher firing rates from dominating the decision.

Testing statistical significance. To assess whether the decoding accuracies we obtained were higher than those expected by chance, we ran a permutation test that consisted of shuffling the labels first and then running the full decoding procedure (Meyers and Kreiman, 2012). This shuffle procedure was repeated 200 times to generate a null distribution, and the results were considered significantly above chance if the real decoding results were higher than all the values in the null distribution ($p < 1/200 = 0.005$). All p values reported in the body of the paper are based on using data in a 300 ms bin that started 100 ms after stimulus onset.

To assess whether there was a statistically significant increase in pose-invariant information from ML/MF to AL, we ran a permutation test (Wasserman, 2004) by combining the data from ML/MF and AL into one population, randomly selected a population of neurons from this joint ML/MF-AL population that was the same size as the original AL population, and calculating the same mean pose-invariant decoding accuracy value that was computed to create Figure 6 (as described below). This procedure was repeated 200 times, selecting a different subset of neurons

each time to create a null distribution under the hypothesis that the ML/MF and AL population had equal pose-invariant information. A *p* value was calculated by assessing how many of the values in the null distribution were as large as or larger than the real pose-invariant decoding accuracy from AL. A similar procedure was used to assess if there was a statistically significant increase in pose-specific information, and to assess whether there was a statistically significant increase in pose-invariant and pose-specific information from AL to AM. Again, we considered results statistically significantly for a *p* value <0.005, and we used data in a 300 ms bin that started 100 ms after stimulus onset.

Face-object-body analyses. Three different decoding analyses were run on data obtained during presentation of FOB stimuli. In the first analysis, we decoded individual images regardless of the category the images belonged to (Fig. 2). A classifier was trained to discriminate between all images in the dataset (since there were 128 images, chance level is 1/128). The firing rates in Figure 2B were calculated by averaging the firing rate over all neurons separately for each image and then sorted based on the decoding accuracy of each image.

In the second analysis, we had the classifier discriminate between exemplars within each category (Fig. 3). To do this analysis, we ran a separate decoding procedure for each category, where the classifier needed to discriminate between the 16 images in a given category (chance 1/16); the results from all eight of these within-category decoding analyses are plotted against each other in Figure 3. We also assessed whether higher population firing rates were positively correlated with higher decoding accuracies using the individual image decoding accuracies that were calculated separately for each category. To do this analysis, for each category, we correlated the population firing and the decoding accuracy for the 16 images in the category. The analysis was repeated for each category and brain region (yielding a total of $3 \times 8 = 24$ correlations). A permutation test was run to see whether the results were statistically significant (one-tailed, $p < 0.05$). We also ran a similar permutation test as a control to see if any of the negative correlations were statistically significant (none were).

In the third analysis, we discriminated faces from nonface objects (Figs. 4A, and 8A) by running seven separate analyses with data from face images in one class, and data from one of the seven nonface object categories in the second class. For each of these seven face versus one other category analyses, two pseudo-population responses of 12 unique face and nonface stimuli were used for training, and testing was done using one pseudo-population response to the remaining four stimuli in each class. (Since we are using different stimuli in the training and test sets, we could have actually trained and tested the classifier on three pseudo-population responses for each stimuli in this analysis, but we decided to use two training responses and one test response to be consistent with all the other analyses in the paper.) This procedure was repeated 50 times using 12 randomly chosen images for training and the remaining four images for testing. The final results plotted in Figures 4A and 8A are averaged over all seven of the face versus object category decoding analyses. To assess whether the decoding performance on this face nonface discrimination task was high, we performed a similar analysis where we had the classifier discriminate between data from nonface category 1 versus a different nonface category 2. A similar procedure was used where the classifier was trained on two examples from 12 randomly chosen stimuli and tested on one example of each of the four remaining stimuli from each class. This led to 21 binary category 1 versus category 2 results, and we plotted the average of these 21 of these binary decoding analyses in Figures 4B. We also plot the results from all 28 of these binary decoding analyses in Figure 4C.

Face-views analyses. Two different decoding analyses were conducted on the face-views dataset to characterize head orientation and face identity coding. In the first analysis (Figs. 5, 6), we assessed pose-invariant identity information by running a series of decoding analyses where we trained a classifier to discriminate between each of the 25 individuals using images from one head orientation, and then we tested the classifier using images from either the same head orientation (using data from different trials) or using images from a different head orientation (thus each bar on each subplot in Fig. 5 comes from running a separate analysis where the classifier was trained and tested on a specific pair of head

orientations). Because the classifier needed to discriminate between 25 individuals, chance is 1/25. The cyan bars in Figure 5 show statistically significant decoding accuracies using a permutation test at the $p < 0.005$ level. Time-resolved results (Fig. 6) were obtained by averaging all the results when the classifier was trained and tested at the same head orientation ("same-pose" decoding accuracy; Fig. 6, blue bars), or averaging all the results when the classifier was trained and tested at different head orientations ("pose-invariant" decoding accuracy; Fig. 6, red bars). Error bars indicate the SDs over all these results. We excluded images from the back of the head from this plot to be sure that they were not driving the differences between AL and AM, although the results look very similar when the back of the head results are included.

For the second analysis, we quantified identity-invariant head orientation information (Fig. 7). To do this we trained a classifier to discriminate between the eight head orientations using two examples for each orientation from 20 randomly chosen individuals, and then tested the classifier using one example of each head orientation from the remaining five individuals (this analysis is similar to the face vs nonface analysis on the FOB where we trained and tested on data from different images so that we could assess information that was invariant to the exact images used). This procedure was repeated 50 times randomly selecting a different 20 individuals to train the classifier each time. Since eight head orientations were used in this analysis, chance was 1/8. For the results plotted as a function of the *k* most selective neurons (Fig. 7B), the top *k* most selective neurons were chosen based on the *p* values from an ANOVA applied to the training data (a separate decoding procedure was run for each value of *k*). The classifier was then trained and tested using only these *k* most selective neurons, and the results are plotted as a function of *k*.

Examining "signal-to-noise" difference between the face patches. To assess whether there were differences between the signal-to-noise ratios (i.e., trial-to-trial variability) between the neurons in different patches that could influence the results, we calculated coefficient of variation values for all neurons used in the decoding analyses using data in a 300 ms bin starting 100 ms after stimulus onset. For each neuron, we took the median CV value across all stimuli as representative of the average level of trial-to-trial variability for that neuron (ignoring stimuli that had produced no spikes). We then compared the distributions of these CV between the neurons in different face patches using a Kruskal–Wallis test, and pairwise Kolmogorov–Smirnov tests. At an α-level of 0.05, none of these tests revealed a statistically significant difference between the patches. We also took the square root of the spike count firing rates on each trial and then redid these analyses and also did not see any statistically significant differences between the patches.

Results

Assessing face and nonface visual information

Our first set of analyses examined how much information the different face patches contained about face and nonface stimuli. While it is clear from previous work that firing rates of neurons are higher for faces than nonface objects in these patches (Tsao et al., 2006; Freiwald and Tsao, 2010; Issa et al., 2013), it is not clear how much information there is about faces compared with nonface objects in each area. In particular, while Tsao et al. (2006) showed that ML/MF can discriminate and categorize nonface objects with above chance accuracy, there has been no systematic analysis of how this evolves along the face-processing pathway, proceeding to more anterior patches. To address this question we applied three different decoding analyses to firing rates from a population of neurons (i.e., population response vectors) that were recorded during presentation of FOB stimuli (the FOB stimulus set contains 16 images from eight different categories; see Materials and Methods). For this and all subsequent analyses, we compared the amount of information in three different face patches: the ML/MF, AL, and AM.

In our first analysis we decoded the identity of each individual image, disregarding category (since there are 128 images in total,
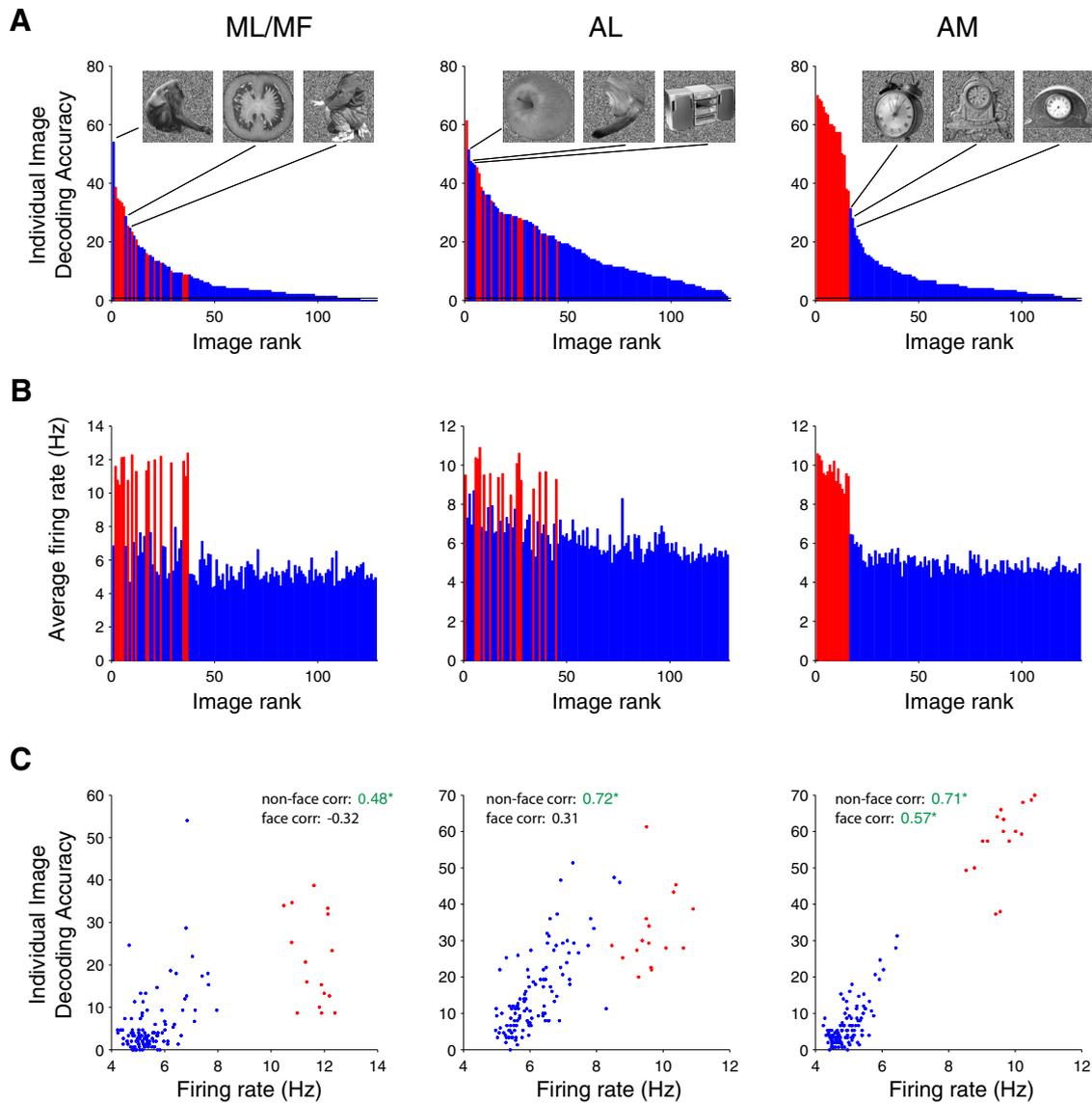
Figure 2. Individual image decoding accuracies and average firing rates in the three face patches to each image in the face, body, and object stimulus set. **A**, The sorted decoding accuracies for each image in the FOB set. Red bars indicate decoding accuracies of face images. The three nonface images with the highest decoding accuracies for each patch are shown as insets at the top of each subplot. The decoding accuracies for all face images are in the top 50% for ML/MF and AL, and all 16 face images had the highest decoding accuracies in AM. **B**, The average population firing rate for each image (in the same order as **A**). Total population firing rates for all face images (apart from one image in AL) were higher than for nonface images. **C**, Individual image decoding accuracies plotted as a function of individual image firing rates. There is a positive correlation between decoding accuracies and firing rates for both face and nonface images (significant correlations are denoted in green with an asterisk).

chance based on naively guessing between all images would be 1/128). This analysis assessed the combined effect of image category and image identity information. If a given face patch represented face images with high fidelity, but did not contain information about nonface objects, then we should be able to decode face images but we should not be able to decode nonface objects. Figure 2A shows the sorted decoding accuracies for each image. The decoding accuracies of face images were generally higher than decoding accuracies of nonface images, which might be expected given the fact that face images elicit much higher firing rates than nonface images and thus could lead to a distinct cluster of face responses (which would make chance decoding for face images closer to 1/16, while chance for nonface images would be around 1/112). However, in ML/MF and in AL several nonface objects were decoded with higher or comparable accuracy as some of the face images, showing that information about nonface

images can be extracted from face patch activity. In ML/MF and only in ML/MF two profile images of crouching bodies were among the best decoded images; in AL there does not appear to be much commonality between the top nonface images; and in AM, the three clock images in the dataset were the nonface images with the highest decoding accuracies. When we plotted the average population firing rate for all images sorted by decoding accuracy (Fig. 2B), we observed that all face images (apart from one in in AL) had a higher population firing rate compared with every nonface image, as was previously shown by Freiwald and Tsao (2010), and is similar to the results found in the human fusiform face area (FFA) by Mur et al. (2012), who showed that almost all individual face images in the FFA elicited higher fMRI responses to face images compared with images from other categories. A scatter plot of firing rate versus decoding accuracy (Fig. 2C) reveals a significant correlation in all three patches for both face and
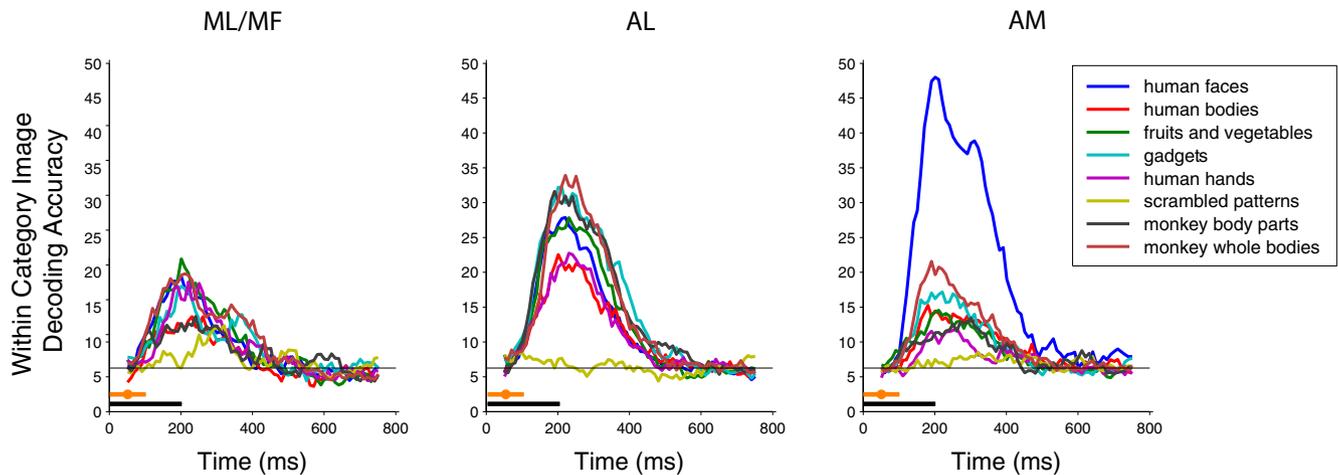
Figure 3. Decoding accuracies for discriminating between the 16 objects within each category on the face, body, and object stimulus set. The within-category decoding accuracy for faces is similar to the accuracy levels seen in several other object categories in ML/MF and AL (right, center subplots) while the within-category decoding accuracy for faces is much higher than the decoding accuracy of other object categories in AM (right subplot). These results show that for most categories, images within the same category elicit different neural responses. The black bars under the plots indicate the time when the stimulus was present, the orange bars under the plot indicate the length of the decoding time window, with the orange circle indicating that the decoding accuracies are aligned to the center of this bin. Chance decoding is 1/16.

nonface objects, suggesting that features decoded by the cells in the nonface objects are ones that tend to produce a higher firing rate, possibly because they share a feature with faces (e.g., the three clocks in AM that showed the highest decoding accuracy among nonface objects also elicited the highest firing rate). The correlation between higher firing rates and decoding accuracies is explored in more detail below.

All patches contain information to discriminate between individual images within each category, with AM showing a very strong representation for specific faces compared with specific objects

In the previous analysis, information about object category (faces, bodies, etc.) could affect the ability to decode individual images. Namely, if the classifier could tell what category an image came from (which might be easier to do for face images given that they all have high firing rates), then it could be easier to tell the identity of an individual. To address this issue, we applied a second decoding analysis where we examined how well we could discriminate between the 16 images within their own category only. This analysis allows us to examine whether all members of a category evoked similar response patterns (e.g., do all face images elicit the same high firing rates across all neurons?), or whether different members within a category evoke distinct patterns of neural activity. Results (Fig. 3) show that the classifier was able to reliably discriminate between members in each category for all of the face patches ($p < 0.005$ permutation test), with the exception of the scrambled image category, which elicited the worst performance in all three patches, and in AL and AM did not exceed our α-level of 0.005 ($p = 0.035$ and $p = 0.060$ for AL and AM, respectively). There were, however, some notable differences between face patches: in ML/MF and AL, it was possible to discriminate between members within each category with a similar level of decoding accuracy for almost all categories, while in AM, the classifier was much better at discriminating between images of faces compared with images in the other categories. Thus we see that ML/MF's and AL's neural activity captures visual features that are present in a range of natural image categories, while again we see that AM represents face images much better than images from other visual categories. It should be noted that the images

Table 2. Correlation between individual image firing rates and decoding accuracies for the three face patches show that higher population firing rates were correlated with higher decoding accuracies

	ML/MF	AL	AM
Human faces	−0.34	0.29	0.50*
Human bodies	0.10	0.62*	0.47
Fruits and vegetables	0.42	0.80*	0.23
Gadgets	0.79*	0.34	0.78*
Human hands	0.60*	0.66*	−0.32
Scrambled patterns	−0.21	−0.01	0.11
Monkey body parts	0.55*	0.89*	0.34
Monkey whole bodies	0.33	0.47*	0.37

Each correlation was calculated separately for the 16 images in each category. Asterisks indicate statistically significant positive correlations ($p < 0.05$) using a permutation test. There were no statistically significant negative correlations.

within particular categories are more visually similar to each other compared with images within other categories, so one should interpret the absolute decoding accuracies for each category within a given face patch with caution. In particular, the face images are more similar to each other in terms of their pixel similarity compared with the other categories, so one might expect lower decoding accuracies for discriminating between different faces compared with other categories based on low-level visual similarity alone (Fig. 1B). Thus one should not read too much into the fact that the decoding accuracies for faces and objects are similar in ML/MF and AL. However, the relationship of decoding accuracies for the different categories across different patches is informative and shows that AM has a much better representation of faces compared with objects relative to the strength of representation between faces and objects in ML/MF and AL.

We also wanted to examine whether images that elicited higher population firing rates also had higher decoding accuracies. To assess whether such a relationship exists, we correlated the individual image decoding accuracies with their population firing rates (this analysis was done separately for each category). Table 2 shows the correlation values for all eight categories for all three face patches. Across the three patches, 10 of 24 ($3 \times 8 = 24$) correlations were significantly positive indicating that there was a
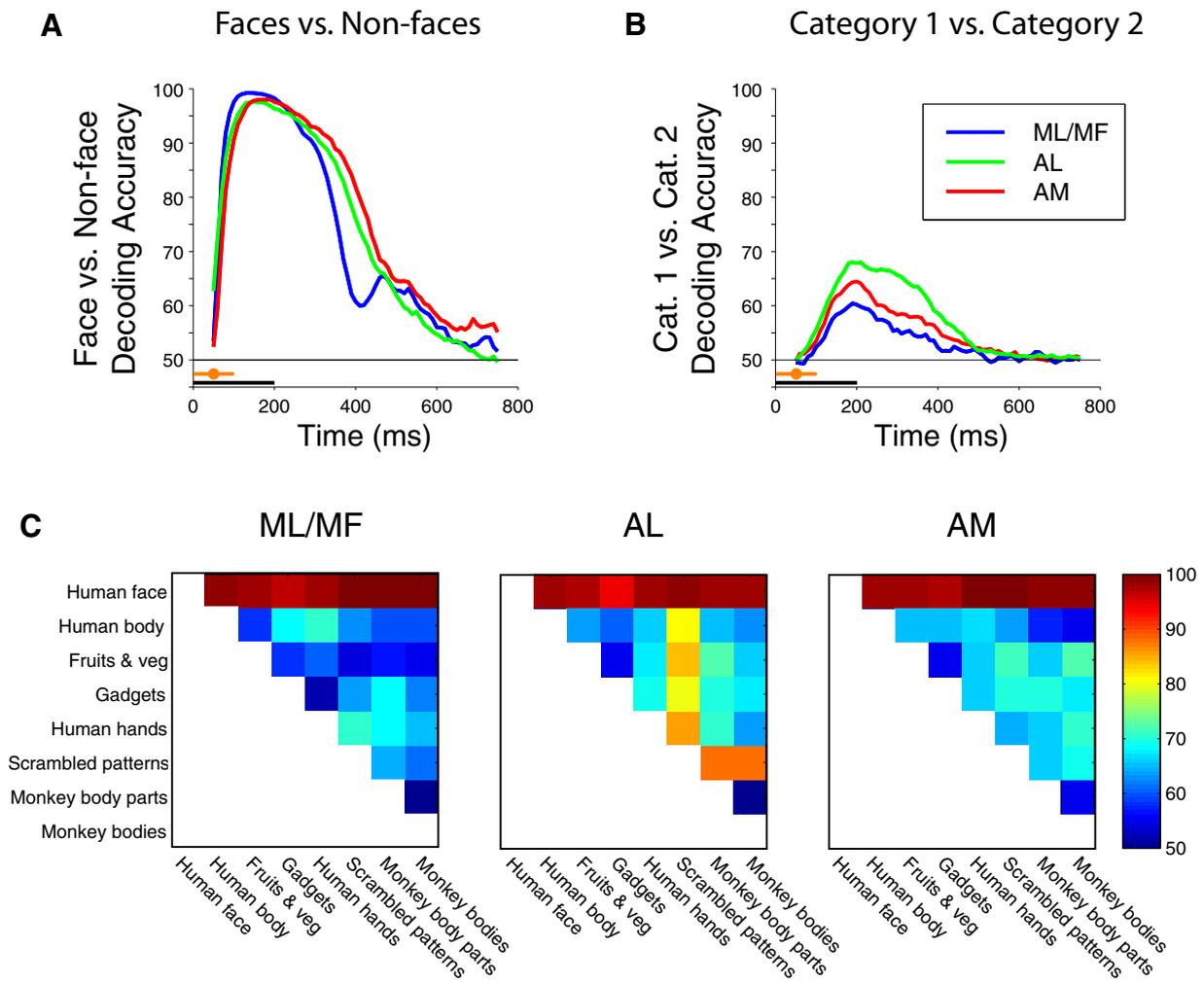
A Faces vs. Non-faces

B Category 1 vs. Category 2

C ML/MF AL AM

Figure 4. Face versus nonface, and category1 versus category2 decoding accuracies. **A**, Decoding accuracies for determining whether an image was a face versus a nonface (face detection) using data from ML/MF (blue), AL (green), and AM (red). The decoding analysis was run separately seven times for face versus each of the nonface categories, and the decoding accuracies were averaged over the seven results. **B**, A comparison of category 1 versus category 2 decoding for the nonface categories (colors for the different brain regions are the same as in **A**). The decoding analysis was run for all 21 pairs of nonface category 1 versus nonface category 2, and the plotted results are the average of each of these 21 decoding analyses. The black bars under the plots indicate the time when the stimulus was present, and the orange bars indicate the length of the decoding time window. **C**, All 28 results for discriminating between all pairs of categories using 300 ms of data (i.e., all pairs of decoding results that went into the averages in **A** and **B**). Chance on these binary decoding tasks is 50%.

relationship between the overall population activity and decoding accuracy (there were no significant negative correlations).

Discrimination performance for faces versus nonfaces is higher than discrimination performance between nonface categories for all patches

Next we examined the capacity of the three face patches for face detection. To address how well populations of cells within the three regions can discriminate between a face from a nonface, regardless of their identity, we set up seven binary decoding tasks, where the exemplars from the first class were always the 16 images of faces, and the examples for the second class were 16 images from one of the nonface categories. For each task, the classifier was trained on 12 randomly chosen faces and nonface images, and tested on the remaining four face and nonface images (the process was repeated 50 times randomly choosing different training and test images each time; see Materials and Methods). Figure 4A shows the results for this face/nonface analysis averaged over all of the seven face versus nonface binary decoding tasks. The results show that all three face patches performed nearly perfectly on this face-detection task. Also, as a reference, we asked whether

the face patches allowed for the discrimination of object categories other than faces. We performed another set of binary decoding analyses trying to decode the 16 images from one of the nonface categories compared with the 16 images from a second nonface category, using the same paradigm of training on 12 images and testing on four images. Figure 4B shows the results for this category discrimination analysis averaged over all of the 21 permutations of the two category decoding tasks. The results show that all face patches perform better than chance in discriminating one nonface object category from another. Since this result was achieved without including the face category, this result is not a simple byproduct of categorizing the response to given stimulus as "not a face"; there must be genuine shape information in the neural population telling one object category from another. Performance on nonface stimulus categorization is much lower than for face detection in all face patches (face vs nonface category = 99, 97, and 99%; nonface category 1 vs nonface category 2 is 61, 70, and 65%, for ML/MF, AL, and AM, respectively). However, this lower performance might in part reflect the lower homogeneity of nonface compared with face images. Among the three areas, AL is best able to discriminate

between different nonface categories. When looking at the results separately for each of the binary category discriminations (Fig. 4C), we see that higher average between category decoding accuracy in AL was due to the fact that it could easily discriminate the scrambled image category from all other categories. This result shows that all 16 scrambled images elicited a similar response pattern in AL (that was different from all other categories) and is consistent with the results in Figure 3, showing that it was hard to discriminate between different members in this category.

Assessing face identity and pose information

The above results show that neural activity in ML/MF, AL, and AM contains information that could be used to detect the presence of a face, and to discriminate among different face and non-face images. For our next set of analyses we sought to look at face representations in more detail and determine the extent to which face patches contain information about identity of a face despite changes in orientation and about the orientation of a face despite changes in identity.

Face identity information across head poses becomes more explicit from posterior to anterior face patches

To examine how much information there was about facial identity regardless of head orientation, we applied a "generalization analysis" (Meyers, 2013) where we trained a classifier to discriminate between each of the 25 individuals using images from one head orientation, and then tested the classifier using images from either the same head orientation using data from different trials (pose-specific identity information) or using images from a different head orientation (pose-invariant identity information). If a neural representation contains face identity information in a pose-invariant format, then decoding performance should generalize across head orientations, even when classifier training is confined to only one orientation. Figure 5 shows the results from training the classifier on images from left profile, straight, or right profile head orientations (Fig. 5, rows) and testing the classifier on all possible head orientations using data from different trials. In ML/MF, significant decoding accuracy is achieved when training and testing the classifier on images of the identical head orientation and, more weakly and in cases only, for physically similar head orientations (straight and upward or straight and half-profile). AL shows generalization of identity information to all head orientations apart from the back of the head, and AM shows even broader generalization that includes even the back of the head. AL and AM also show elevated mirror symmetric decoding accuracies (e.g., high performance for training on the left and testing on the right profile.

To further compare pose-specific identity information to pose-invariant identity information, we averaged all the results that were created from training and testing the classifier at the same head orientation, and we compared them to the average results when the classifier was trained on one head orientation and tested on a different orientation (in both cases we excluded results from the back of the head images to make sure that this condition was not driving any differences seen between AL and AM, although similar results were obtained when the back of the head images were included). Fig. 6 shows the results from this analysis plotted as a function of time. The results show that AL and AM have similar levels of pose-specific identity information ($p = 0.075$), which is higher than the pose-specific identity information in ML/MF ($p < 0.005$), and that there is a clear increase in pose-invariant information from ML/MF to AL to AM ($p < 0.005$ from ML/MF to AL, and from AL to AM), which is

consistent with the notion that these regions form a hierarchy for constructing pose-invariant representations.

Head pose information decreases from posterior to anterior face patches

To compare how much information was present about head orientation that is separate from information that is present about specific individuals, we applied a decoding analysis where we trained a classifier to discriminate between the eight head orientations using two training examples taken from 20 randomly chosen people, and we tested the ability of the classifier to categorize one example from the remaining five people (this procedure was repeated 50 times randomly selecting 20 people to train the classifier on each time, and the results were averaged over these 50 runs). We could best decode head orientation information from AL and ML/MF, with AM showing a lower decoding accuracy (Fig. 7). Thus it appears that AM is building pose-invariant representations of face identity at the cost of losing information about exact head orientation.

To ensure that this decrease in pose information in AM was not somehow due to the classifier putting too much emphasis on the pose-invariant neurons, we did an additional analysis where we used an ANOVA to find the most pose-selective neurons on the training set, and then we trained and tested the classifier using only these selective neurons. Figure 7B shows the decoding results from this analysis as a function of the number of neurons we selected. The results show that even when only a small number of the most selective neurons are used (~3–5), there is less pose information in AM than ML/MF and AM. Thus while AM does possess some neurons that are sensitive to pose, as is evident from the pose decoding accuracy being well above chance, it is clear that this pose information in AM is weaker in even the most pose-selective neurons compared with ML/MF and AL.

Examining coding of information in high population firing rates versus patterns across neurons

Almost all the information present is contained in the pattern of responses across neurons

For our final analysis, we examined whether information is coded in the overall level of activity in the population (i.e., average firing rate over all neurons), or whether information is mainly coded in the differential pattern of activity across neurons (i.e., differential patterns across neurons ignoring the overall population activity). This question is of interest because, as shown in Figure 2C and Table 2, overall population firing rates and decoding accuracy are positively correlated. Thus it is possible that gradations in the overall population firing rate might contain the majority of information for discriminating between stimuli. Furthermore, fMRI recordings appear to be best correlated with the average activity over a larger regions of cortex (Issa et al., 2013), so comparing a classifier that can average all spiking activity to a classifier that takes into account individual contributions of neurons can give insight into what information might be difficult to extract from fMRI signals.

To examine this question, we compared the decoding accuracies of two classifiers that only take the relative patterns across neurons into account (the Maximum Correlation Coefficient classifier and the Minimum Angle classifier), to a classifier that only takes the overall level of population activity into account (the Total Activity classifier), and the PNB classifier (used in all other analyses in the paper), which uses both the overall population activity as well as the relative patterns across neurons. Figure 8A shows the results comparing these four classifiers on the face versus other object category decoding task (using the FOB data-
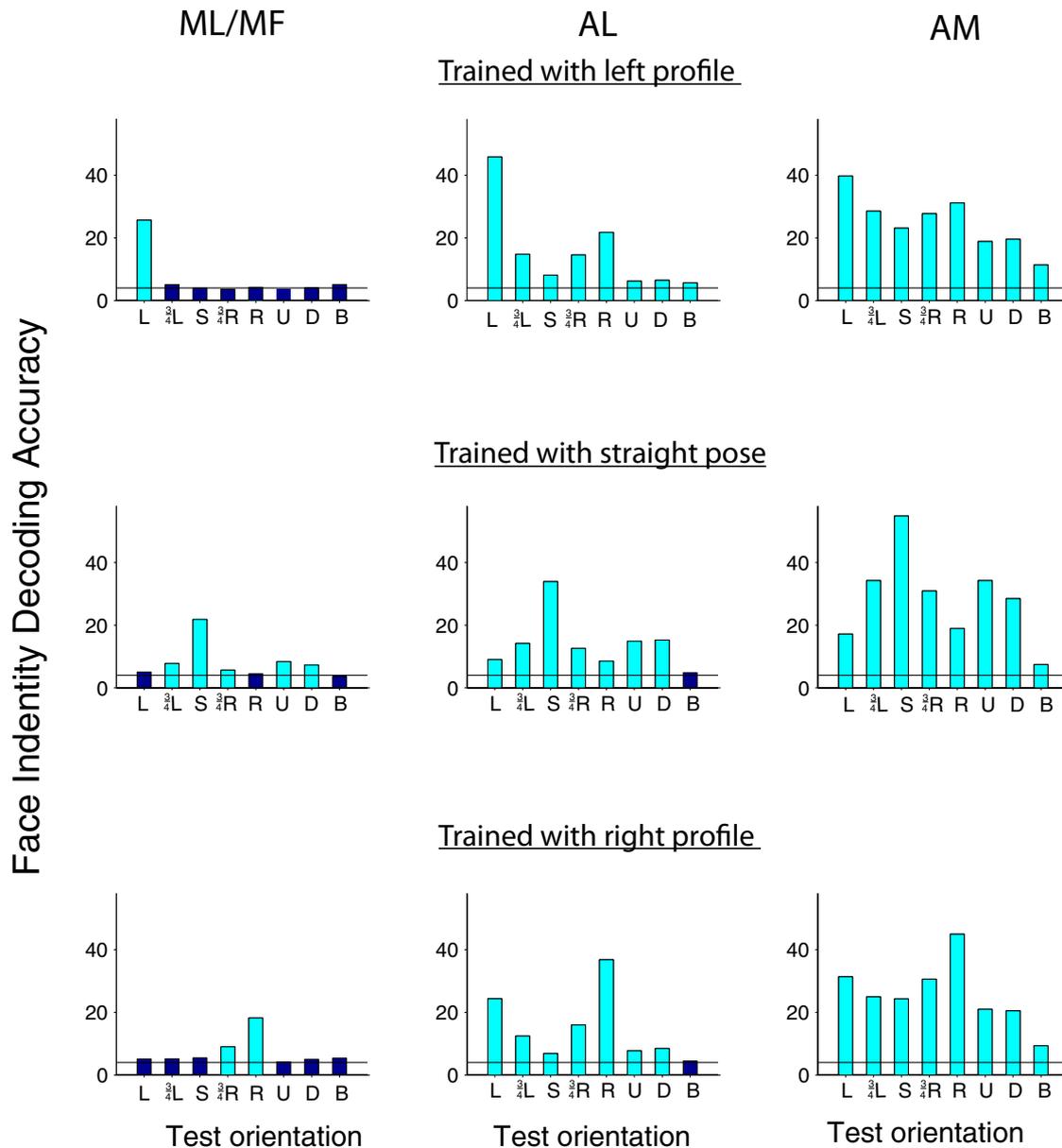
Figure 5. Pose-specific and pose-invariant face identity information. A classifier was trained to discriminate between individuals using data from one head orientation and then tested with data from another head orientation. The left column shows results from ML/MF, middle column from AL, and right column from AM. First row is for training on the left profile, middle row for training on the straight pose, and bottom row for training on the right profile face. Each bar in the histogram shows decoding accuracy as the result of testing at a specific location (L, left profile; 3/4 L, left 3/4 profile; S, straight; 3/4 R, right 3/4 profile; R, right profile; U, head tilted up; D, head tilted down; B, back of the head). Cyan indicates statistical significant results ($p < 0.005$ permutation test), and dark blue indicates failure to reach significance. We see higher decoding accuracy in AL and AM compared with ML/MF. Additionally, we see an increase in pose-invariant generalization from training on one head orientation and testing on a different head orientation going from ML/MF to AL to AM.

set). Performance of all classifiers in the face versus object task was similar, indicating that there is information to discriminate between faces and other object categories based on either the total level of activity in the population or based on the pattern across neurons disregarding the overall level of population activity. In contrast, on all other decoding tasks, we found the Total Activity classifier had a much lower decoding accuracy than the other classifiers. For example, within-category discrimination decoding of the monkey whole bodies on the FOB dataset (Fig. 8B) and pose-specific face identity decoding on the FV dataset (Fig. 8C) were much lower for the Total Activity classifier compared with the other classifiers. Thus, while an overall high population firing rate might indicate the presence of a face, it appears that the overall activity does not have information much beyond that.

Additionally, examining the relative pattern of firing rates across neurons while discounting the overall firing rate (Maximum Correlation Coefficient and Minimum Angle classifiers) achieved equivalent decoding accuracies as using a classifier that took into account both the pattern across neurons and the total activity level (i.e., the PNB classifier), indicating that all the information that we were able to extract was contained in the relative patterns of activity across neurons.

However, this is not to say that there is no relationship between decoding accuracy and higher population firing rates. Indeed, when we correlate individual image firing rates with individual image decoding accuracies, we see a statistically significant positive correlation for both the 128 individual image decoding analysis (Fig. 2C) and when decoding individual images
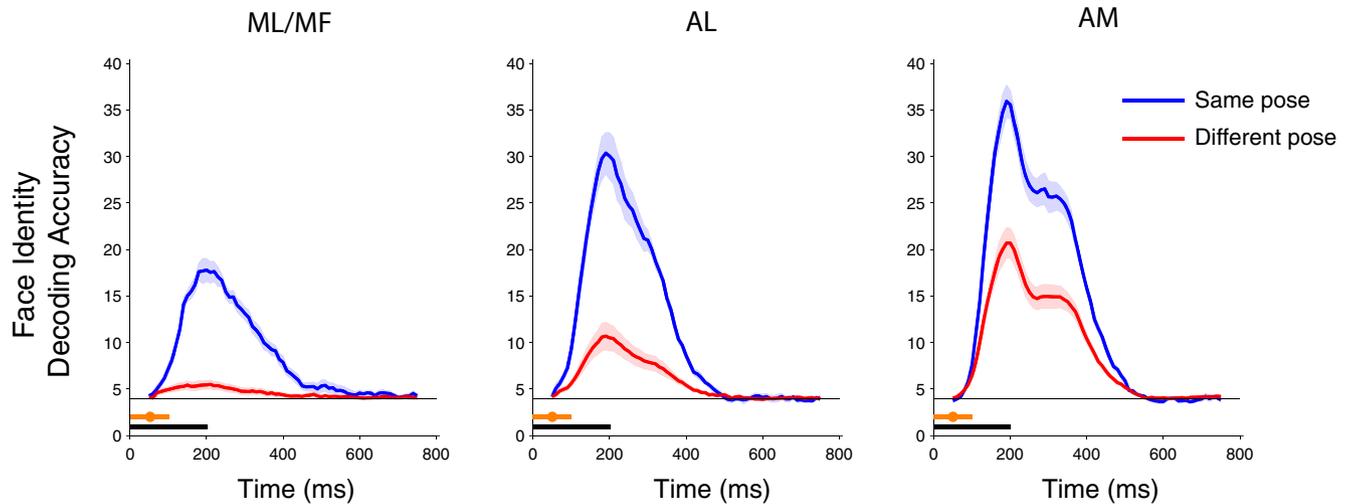
Figure 6. Average pose-specific and pose-invariant identity decoding results. Results from decoding the identity of a face when training and testing a classifier with images from the same head orientation (pose-specific decoding, blue trace) or with images from different head orientations (pose-invariant decoding, red trace). Pose-specific decoding results are averaged over all seven conditions for training and testing on the same head orientation, while the pose-invariant decoding is averaged over all 36 pose-invariant decoding permutations (back of head conditions were not included in this analysis). Error bars indicate the SDs over all these decoding results. The results are shown for ML/MF (left), AL (middle), and AM (right). AL and AM have higher pose-specific decoding accuracies than ML/MF and there is an increase in pose-invariant information from ML/MF to AL to AM. The black bars under the plots indicate the time when the stimulus was present, and the orange bars under the plot indicate the length of the decoding time window.
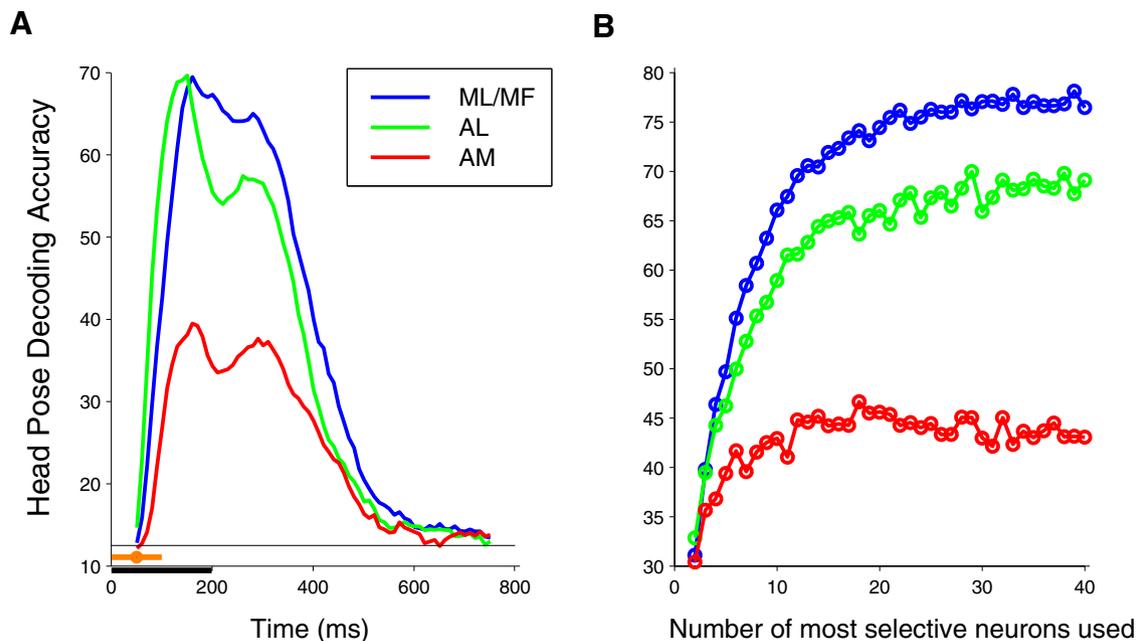

Figure 7. Head pose decoding results. Results from decoding the pose of the head generalizing over facial identity. **A**, Decoding results for the three face patches plotted as a function of time. Head pose is better represented in ML/MF and AL than in AM. The black bars under the plots indicate the time when the stimulus was present, and the orange bars under the plot indicate the length of the decoding time window. **B**, Decoding results plotted as a function of the number of selective neurons used (i.e., results are plotted using the 2 most selective neurons for decoding, the 3 most selective neurons, etc.). The selectivity of neurons was based on ANOVA p values using data from the training set. The results show that AM has lower decoding performance than ML/MF and AL even when only a small number of neurons is used indicated that even the most pose-selective neurons in AM are less pose selective than the most pose-selective neurons in ML/MF and AL.
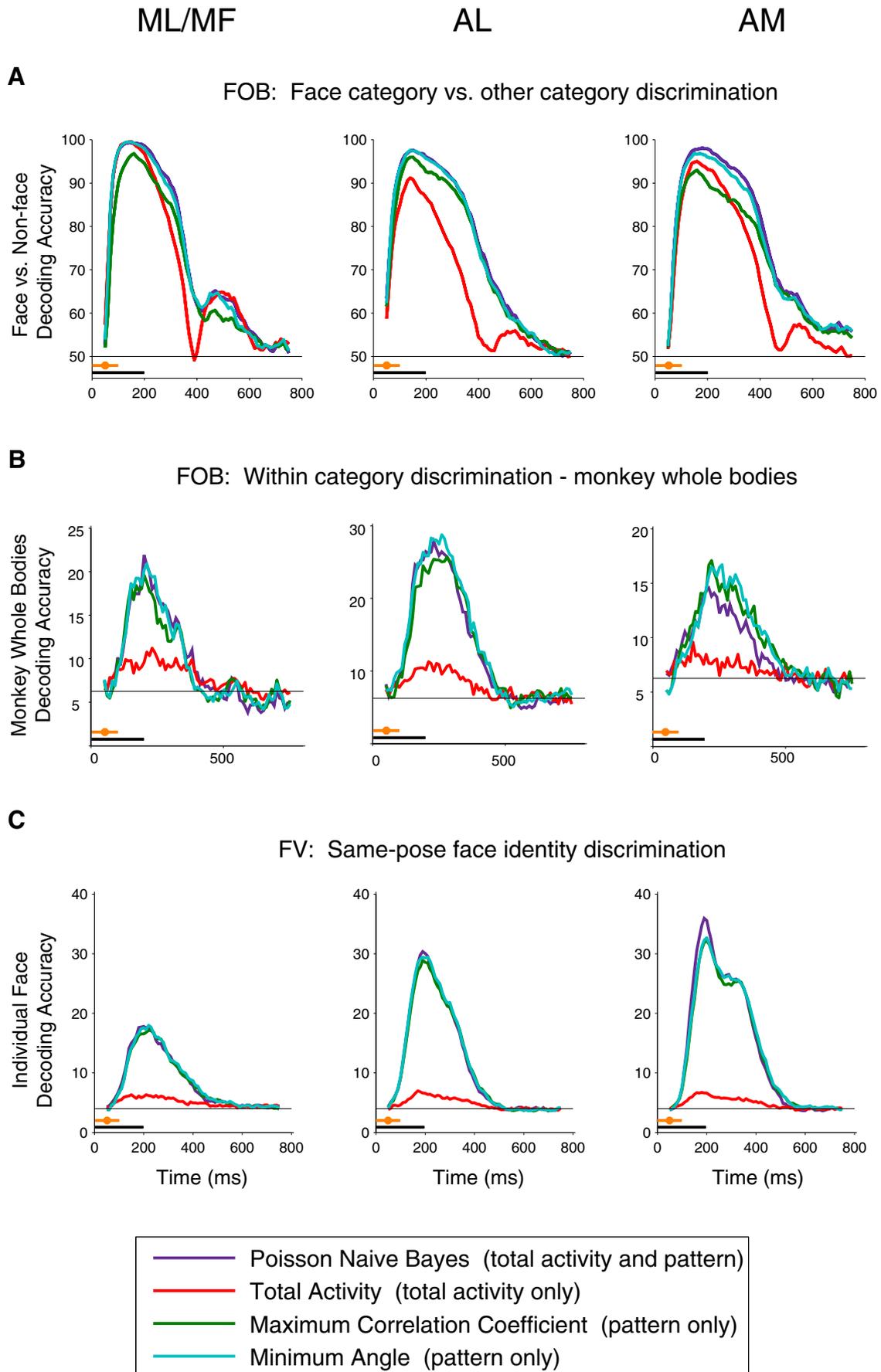
within each category for many categories in ML/MF, AL, and AM (Table 2). Thus higher population activity generally indicates higher decoding accuracy, although a more detailed assessment about what types of information are present cannot be extracted from the more crude measure of total population firing rate.

Discussion

In this study, we exploit the macaque face patch system to address a question of great importance to systems neuroscience: How are population codes hierarchically transformed in the brain? The macaque face patch system provides an ideal opportunity to address this question since it is composed of discrete, functionally distinct, but strongly connected nodes.

We use a population decoding approach to compare how three different face patches (ML/MF, AL, and AM) represent different types of face and nonface stimuli. Our quantitative approach reveals several important new insights. First, we find that we can extract information about both faces and nonface objects from all three patches with above chance accuracy, and impor-

tantly, the AM population appears to selectively represent individual face identity over nonface object identity compared with ML/MF and AL populations (Figs. 2, 3). Second, we found that pose-invariant face identity information appears to increase as one moves to more anterior patches, while information about the orientation of the head regardless of the individual is best represented in the two more posterior patches ML/MF and AL (Figs. 5–7). Finally, we found that much more information to distinguish faces and objects is contained in patterns of activity across neurons than in the total level of population activity (Fig. 8).

Information is contained in patterns of activity in both high and low population firing rate regimes

While previous results showed that neurons in these fMRI-defined face patches respond with higher firing rates to face images compared with nonface images (Tsao et al., 2006; Freiwald and Tsao, 2010; Issa et al., 2013), they did not examine in detail how much information is present in the population about nonface objects. By applying a decoding analysis we are able to see that there is significant nonface information in the lower firing rate regimes in ML/MF and AL, while in comparison, AM had much more information about faces compared with nonface object categories (Figs. 2, 3).

The fact that we could decode nonface information from some of these face patches is similar to the results of Haxby et al. (2001), who showed that it is possible to decode object information from the FFA based on fMRI BOLD responses. However, because it is still not known exactly how information is coded in neural activity, we remain agnostic in our interpretation as to whether the nonface object information in these lower firing rates contribute to behaviors related to nonface objects. For example, particular visual properties of the nonface objects might resemble face-like features,which could allow us to decode information about the nonface objects; however, this could be irrelevant for the way monkeys actually solve the task. While it could also be argued that the converse is true, namely that the face patches might be responding just to visual shapes that faces happen to contain, other evidence makes this less likely (e.g., the contrast preference of face cells seen in Ohayon et al. (2012)). Additional studies that perturb particular face patch regions while monkeys engage in face and object discrimination tasks could help resolve the question about whether this nonface object information in lower firing rates actually contributes to behavior.

Relatedly, our analyses comparing different classifiers (Fig. 8) show that the total population activity level contains relatively little information (apart from information about whether a face was present). Consistent with this, multiple fMRI studies have

demonstrated that distributed response patterns within a brain region contain a rich amount of information above and beyond what is present in univariate modulation (Kamitani and Tong, 2005; Kriegeskorte and Bandettini, 2007; Çukur et al., 2013). Yet given that that higher firing rates are correlated with higher neural decoding accuracies (Fig. 2C, Table 2), even coarse overall activity levels can, at least sometimes, indicate the types of information represented, which could explain why region of interest analyses are useful for inferring the function of brain regions. We believe the link between higher population activity levels and higher decoding is likely due to higher firing rate regimes enabling neurons to code information in larger dynamic range of activation and, consequently, giving rise to more discriminative patterns of activity.

The face representations in AM contain information in a highly invariant manner

Our analyses of the face-views dataset give a more detailed picture of what types of face information are contained in face patch regions. In agreement with Freiwald and Tsao (2010) we see an increase in pose-invariant identity information from ML/MF to AL to AM (Figs. 5, 6), that the neural populations have similar responses to mirror symmetric head orientations in AL and AM (Fig. 5), and that there is more head orientation information in ML/MF and AL than in AM (Fig. 7). Our analyses also reveal that the classifier was able to partially generalize from the back of the head to frontal views using data from AM and that this was not possible using data from ML/MF (Fig. 5), which indicates that global features such as hair, head shape, or skin tone are used by AM; however, generalization was much better to nonbackward views, indicating that view-invariant tuning in AM is not due solely to global cues. Interestingly, monkeys had not interacted with 21 of the 25 individuals in the image set. It thus appears that the pose-invariant identity representations we report here are not a result of extensive familiarity with the individual in the image set. Our results bear some similarity with two fMRI studies. Kriegeskorte et al. (2007) showed that multivoxel activity patterns in human anterior inferotemporal cortex, but not in the more posterior FFA, can be used to discriminate between individual face images, and Anzellotti et al. (2014) decoded face identity information invariant to head orientation from occipital face area, FFA, and inferotemporal cortex.

In previous work we have found that it is useful to view the ventral visual pathway as a system that performs "intelligent information loss," where information about visual details in early visual areas, such as V1, are lost to build up abstract/invariant representations that are more useful for behavior in higher brain regions (Meyers and Kreiman, 2012; Meyers, 2013). The results in this study fit well into this framework where we see a gradual buildup of representations that are better suited for processing faces in AM at the cost of losing information about nonface stimuli (also see Anzellotti et al., 2014). The representations in AM do not appear to be as fully abstract as the neurons found in the human medial temporal lobe, which respond selectively to a particular concept of a person even when the visual image eliciting neural responses have nothing in common (such as the written name and an image of a particular person; Quiroga et al., 2005).

Outlook

While the results presented here further our understanding of the face patch system, they also help to highlight the importance of future studies that would be useful for gaining a deeper understanding of how this system is operating. In particular, the fact

←

Figure 8. Classifier comparison to assess how much information is in the overall firing rate level and in patterns across neurons. Each subplot compares the results from four different classifiers (Maximum Correlation Coefficient classifier in green, Minimum Angle classifier in cyan, Poisson Naive Bayes classifier in purple, and Total Activity classifier in red). The columns contain data from the three different face patches. *A*, Results from decoding whether an image was a face or is in another object category on the FOB dataset (i.e., same decoding task as in Fig. 4). *B*, Within-category discrimination for monkey whole bodies on the FOB dataset (i.e., same decoding task as in Fig. 3). *C*, Pose-specific individual face discrimination on the FV dataset (i.e., same as the blue traces on Fig. 6). For face versus nonface results, all classifiers performed similarly. For all other analyses, there was much more information in classifiers that extracted patterns of information across neurons (PNB, Maximum Correlation Coefficient, Minimum Angle) than the classifier that used the total activity (TAct). The black bars under the plots indicate the time when the stimulus was present, and the orange bars under the plot indicate the length of the decoding time window.

that the complexity of face-specific features appears to be building up in the face patch system has shown us that it would be useful to obtain recordings from more neurons using an even larger set of visual images. A study of the posterior face patch (PL) by Issa and DiCarlo (2012) has given strong evidence that the early responses to this most posterior face patch are largely driven by an image of an eye and an outline of the face, and recordings by Freiwald and Tsao (2010) in the middle face patches show that these patches are driven by features relating to the eyes (iris size and intereye distance) and overall face shape (aspect ratio and face direction; Freiwald et al., 2009). By using a broader stimulus set and systematically comparing across posterior, middle, and anterior patch, it should be possible to gain a better understanding of the visual features that drive the populations of neurons recorded from in this study, and potentially relate these features to computational models (Leibo et al., 2014; Tan and Poggio, 2014). Additionally, while the monkeys in this study did have a lifetime of exposure to human faces, it would be useful to examine how the face patches respond to more ecologically relevant macaque face images, in addition to the human face images used here, to determine whether the results hold across species. Overall, however, the results shown here underscore the functional differences between face patches, and help set a foundation for understanding what information is read out about faces and non-face objects from each patch, by quantifying what information is available.

Notes

Supplemental material for this article is available at figshare. com/s/8bfec6b8cce811e4866406ec4bbcf141. The supplemental file contains additional analyses of the data. This material has not been peer reviewed.

References

Anzellotti S, Fairhall SL, Caramazza A (2014) Decoding representations of face identity that are tolerant to rotation. Cereb Cortex 24:1988–1995. CrossRef Medline

Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77: 305–327. CrossRef Medline

Çukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. Nat Neurosci 16:763–770. CrossRef Medline

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051–2062. Medline

Duda RO, Hart PE, Stork DG (2001) Pattern classification. New York: Wiley.

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

Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. Nat Neurosci 12:1187–1196. CrossRef Medline

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

Issa EB, DiCarlo JJ (2012) Precedence of the eye region in neural processing of faces. J Neurosci 32:16666–16682. CrossRef Medline

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

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

Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol 71:856–867. Medline

Kriegeskorte N, Bandettini P (2007) Analyzing for information, not activation, to exploit high-resolution fMRI. Neuroimage 38:649–662. 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

Leibo JZ, Liao Q, Anselmi F, Poggio T (2014) The invariance hypothesis implies domain-specific regions in visual cortex. Cold Spring Harbor Laboratory Journals. doi: 10.1101/004473.

Meyers EM (2013) The neural decoding toolbox. Front Neuroinform 7:8. CrossRef Medline

Meyers EM, Kreiman G (2012) Tutorial on pattern classification in cell recording. In: Visual population codes (Kriegeskorte N, Kreiman G, eds), pp 517–538. Boston: MIT.

Moeller S, Freiwald WA, Tsao DY (2008) Patches with links: a unified system for processing faces in the macaque temporal lobe. Science 320:1355–1359. CrossRef Medline

Mur M, Ruff DA, Bodurka J, De Weerd P, Bandettini PA, Kriegeskorte N (2012) Categorical, yet graded–single-image activation profiles of human category-selective cortical regions. J Neurosci 32:8649–8662. CrossRef Medline

Mutch J., Knoblich U., Poggio T (2010) CNS: a GPU-based framework for simulating cortically-organized networks. Technical Report. MIT-CSIAIL-TR-2010–013/CBCL-286. Cambridge, MA: MIT.

Ohayon S, Freiwald WA, Tsao DY (2012) What makes a cell face selective? The importance of contrast. Neuron 74:567–581. CrossRef Medline

Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329–342. CrossRef Medline

Pokorny JJ, de Waal FB (2009) Face recognition in capuchin monkeys (Cebus apella). J Comp Psychol 123:151–160. CrossRef Medline

Quian Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 10:173–185. CrossRef Medline

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. Nature 435:1102–1107. 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

Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci U S A 104:6424–6429. CrossRef Medline

Tan C, Poggio T (2014) Neural tuning size is a key factor underlying holistic face processing. CBMM Memo No. 21.

Tsao D (2006) Eppendorf 2006 winner. A dedicated system for processing faces. Science 314:72–73. 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

Wasserman L (2004) All of statistics. New York: Springer.