A Cortical Region Consisting Entirely of Face-Selective Cells

Doris Y. Tsao[1,3,*,†], **Winrich A. Freiwald**[1,3,*], **Roger B. H. Tootell**[2,4,5], and **Margaret S. Livingstone**[1]

1*Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA.*

2*Department of Radiology, Harvard Medical School, Boston, MA 02115, USA.*

3*Brain Research Institute, Center for Advanced Imaging, University of Bremen D28334, Bremen, FR Germany.*

4*Massachusetts General Hospital, Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA 02129, USA.*

5*Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.*

Abstract

Face perception is a skill crucial to primates. In both humans and macaque monkeys, functional magnetic resonance imaging (fMRI) reveals a system of cortical regions that show increased blood flow when the subject views images of faces, compared with images of objects. However, the stimulus selectivity of single neurons within these fMRI-identified regions has not been studied. We used fMRI to identify and target the largest face-selective region in two macaques for single-unit recording. Almost all (97%) of the visually responsive neurons in this region were strongly face selective, indicating that a dedicated cortical area exists to support face processing in the macaque.

Lesion studies show that object recognition depends on the temporal lobe (1), but the principles of temporal lobe organization underlying the representation of objects remain uncertain. In particular, the question of how face processing is functionally organized has been a focus of intense debate (2–4). In humans, several cortical regions have consistently been found in fMRI studies to be more responsive to faces than to other objects, and it has been suggested that the fusiform face area (FFA) is exclusively dedicated to face processing (5). However, physiologists who are recording from the macaque temporal lobe have never found any entirely face-selective region; instead, they have reported scattered clusters of face-selective cells, especially prevalent in the upper and lower banks of the superior temporal sulcus (STS), with, at most, 20 to 30% of the cells in any region being face selective (6–9).

It is possible that an area consisting entirely of face-selective cells exists in the macaque and has simply been missed because of single-unit sampling limitations. Alternatively, no such

---

area may exist, and regions of the macaque brain identified by fMRI as face-selective (10, 11) may actually contain a mixture of cells selective for both faces and nonface objects. fMRI measures average blood flow within sampling units containing hundreds of thousands of cells, and therefore it cannot directly address the selectivity of single units. To clarify the neural organization of face processing, we used fMRI to target single-unit recordings to the middle macaque face patch. Our goal was to understand the selectivity of single neurons within this specific ~16-mm^2 (12) region of the temporal lobe, which appears to be topographically homologous to the human FFA (10).

Single-unit recordings targeted to fMRI-identified face-selective regions were performed in two monkeys, M1 and M2, in a standard electrophysiology setup outside the scanner. We first localized the face-selective regions in both monkeys with fMRI (Fig. 1, A and B), and we then implanted a recording cylinder roughly over the targeted region. A second anatomical scan with magnetic resonance (MR)–visible markers in a grid inside the recording cylinder showed precisely which grid holes in the chamber targeted the center of the face patch (Fig. 1C). Recordings were made from three adjacent grid-hole positions in both monkeys. A guide tube was placed in the grid hole to allow reliable access to the face patch.

We used fMRI to identify face patches. Ninety-six images of faces, bodies, fruits, technological gadgets, hands, and grid scrambled patterns (16 images per category, one category per block; see fig. S1 for example stimuli) were presented to the monkey during continuous central fixation. To optimize the signal-to-noise ratio, we used the exogenous iron oxide contrast agent MION (monocrystalline iron oxide nanoparticle) (13). Consistent with previous results (10, 11), in both monkeys several discrete regions (face patches) responded significantly more to faces than to five other object categories. The most prominent face patch in both monkeys was the one located at A6 (i.e., 6 mm anterior to the interaural line) (Fig. 1, A and B). In addition, monkey M1 had a more posterior face patch located at A0, and both monkeys had anterior face patches located between A15 and A22 (fig. S2). We designate the patch located at A6 the "middle face patch" throughout this paper, to distinguish it from the anterior face patches and from the region posterior to A6, which showed variable face selectivity across monkeys.

Figure 1A shows a semisagittal section from monkey M1 in which all three face patches are visible. We targeted the middle face patch for single-unit recordings because it was the most prominent in both monkeys (and in all the monkeys we have scanned so far, $n = 7$ monkeys) and because of its possible homology to the human FFA (10). In monkey M1, this patch was located on the lip of the lower bank of the STS; whereas in monkey M2, it was located in the fundus of the STS (Fig. 1B). This individual difference underscores the importance of using fMRI to target single-unit recordings in the same animal. The lesion left by the recording guide tube in monkey M1 is visible in Fig. 1, A and C, and confirms, in three dimensions, that our single-unit recordings accurately and precisely targeted the middle face patch.

We tested the face selectivity of 405 single units (241 in the right hemisphere of monkey M1 and 164 in the left hemisphere of monkey M2) in the middle face patch with the same 96 images used to localize the face patches with fMRI. The stimuli were presented foveally every 400 ms (200 ms on and 200 ms off) in random order for 4 to 10 repetitions while the monkey fixated. We recorded responses from all single units encountered, regardless of visual responsiveness or face selectivity. Across the population of recorded cells, 182 of 241 (76%) cells in monkey M1 and 138 of 164 cells (84%) in monkey M2 gave significant responses (14) to at least one of the 96 images and were therefore classed as visually responsive.

Figure 2A shows the normalized response selectivity of all visually responsive cells recorded from the two monkeys (15). Each of the faces (images 1 to 16) elicited stronger responses across the population than did any of the 80 nonface objects (images 17 to 96). Figure 2B shows

bar graphs of the average responses to each of the 96 images across the population of visually responsive cells. In monkey M1, the ratio of face to nonface object response was –74 (negative, due to a small suppression to the nonface objects on average); in monkey M2, this ratio was 21.

In addition to the overwhelming bias for face stimuli, many cells gave significant responses to a few particular nonface objects (the faint orange lines in Fig. 2A to the right of the first 16 columns). In monkey M1, the two nonface objects that gave mean responses across the population exceeding six average standard errors (SEs) were a clock and an apple (Fig. 2C). In monkey M2, the only nonface objects that elicited significant responses across the population were also round. The small but significant responses to round stimuli suggest that the coding of faces in the middle face patch is based on analysis of visual shape.

To quantify the face selectivity of individual cells, we defined a face-selectivity index as FSI = (mean response$_{faces}$ – mean response$_{nonface\ objects}$)/(mean response$_{faces}$ + mean response$_{nonface\ objects}$). Figure 2D shows population histograms of the FSI in both monkeys. The distributions are strongly skewed toward high FSI values. The mean absolute magnitude of the FSI was 0.90 in monkey M1 and 0.87 in monkey M2, which correspond, respectively, to a 19:1 and a 14:1 ratio of face-to-nonface object response. In the single-unit literature, cells are typically classified as face selective if they respond at least twice as strongly to faces as to nonface objects (16, 17). By this criterion, all but 8 out of 310 total visually responsive cells, or 97%, were face selective (we considered cells that were selectively inhibited by faces to be face selective as well; if we required an excitatory response to faces, then 280/310 = 90% of cells were face selective).

Because the monkeys were highly familiar with the 16 screening faces, one could hypothesize that exposure to these specific faces contributed to the cells' selectivity. This appears unlikely for two reasons. First, we found that the face selectivity of units in the middle face patch did not depend on the particular set of images tested. Although some cells responded best to only one or a few faces (Fig. 2A, left, cell 40), many cells were responsive to a wide variety of face images, including familiar and unfamiliar faces, human and macaque faces, and even cartoon faces (fig. S3). Cells maintained their face selectivity when tested with a wide variety of novel face and nonface images, including monkey headless bodies and body parts, as well as hundreds of natural images. Second, because each of the 96 images in the screening set was shown equally often, we do not believe that selectivity for one particular subset of images (faces) could emerge from repetitive passive viewing of the whole set of images.

What was the selectivity of units that did not give a clear response to any of the 96 screening stimuli? When we encountered such a unit, in most cases we documented the nonresponsiveness and then advanced the electrode in search of the next unit. However, in cases where we were recording from a pair of units simultaneously and only one was visually responsive, we tested a battery of additional face stimuli (18) on the nonresponsive unit as well. We found that out of 14 initially non–visually-responsive units tested in this way, 9 actually were responsive to face stimuli but were selective for nonfrontal views, different expressions, or monkey faces. None of the remaining five units showed a significant response to any nonface object. It is therefore likely that many, if not all, non–visually-responsive units were similarly selective for face characteristics not included in the set of screening stimuli.

The local field potential (LFP) represents summated excitatory and inhibitory synaptic potentials in thousands of neurons around the electrode tip. It has been reported that the LFP correlates better than single units with the fMRI signal (19). Evoked LFPs recorded from monkeys M1 and M2 are shown in fig. S4. In both monkeys, two large face-selective troughs with peak magnitudes at 130 ms and 200 ms were evident in the LFP. We observed these face-

selective LFP troughs at almost all recording sites in the middle face patch, providing further evidence that population activity within this face patch was strongly face selective. The existence of two face-selective troughs suggests two discrete stages of face processing, possibly triggered by the arrival of feedforward and feedback/recurrent inputs, respectively.

One fundamental function of face processing is to identify individuals. Cells responding sparsely and robustly can be used not just to detect the presence of a face, but to discriminate the identity of a particular face. To measure how much information these face-selective cells carried about face identity (i.e., differences between different face images), we examined all cells for which the 96 faces and objects had been presented at least five times (94 cells). The response magnitude elicited by the 96 images in these 94 cells on four trials was averaged to yield, for each image, a population vector ("template vector"). We then asked whether we could predict the identity of an unknown image from the activity it elicited across the population of 94 cells on the remaining fifth trial ("test vector"). Identification was performed by determining the template vector to which the test vector was closest in Euclidean distance (Fig. 3A). If the population response correctly identified a particular image, then the 96×96 matrix of (test vector, template vector) distances should have a minimum value on the diagonal in the row corresponding to that image (chance = 1/96). We also examined categorization by using the following test: If the population response correctly categorized a particular image as belonging to one of the six stimulus categories, then the response to that image should be closest to the mean of the 16 template vectors in the same category (chance = 1/6).

Figure 3B shows the percent correct identification and categorization obtained using this algorithm. Mean individual face identification accuracy was 74%, and mean face categorization accuracy was 100%. Thus, information about face category and identity is available within this patch of cortex. Performance was significantly better for faces than for nonface objects, for both individual identification (t test, $P < 2 \times 10^{-16}$) and categorization (t test, $P < 5 \times 10^{-19}$).

Evidence from human psychophysics suggests that objects are identified at the category level (e.g., face versus fruit) before they are identified at the individual level (20). A physiological correlate of this is that information about face category precedes information about face identity by an average of 51 ms in face-selective cells recorded from the anterior STS (21,22). To examine the time course of identity and category information in single units in the middle face patch, we computed time-varying test and template vectors from average responses within a 50-ms sliding window. Categorization performance reached its maximum earlier (133 ms) than did identification performance (192 ms) (Fig. 3C).

Single units in the middle face patch of the macaque temporal lobe showed a remarkable specificity for face processing and contained a much higher concentration of face-selective cells than reported previously. Indeed, the only nonface images that elicited significant (albeit small) responses across the population were clocks and round fruits, which share a common shape attribute with faces. In agreement with previous single-unit studies of face-selective cells in the temporal lobe, cells in the middle face patch carried information about the identity of individual faces distributed across the population (23–25), and they showed a face-inversion effect (17,26). The responses of cells in the middle face patch clearly tended toward distributed coding within the domain of faces, because many cells were activated by a wide range of face stimuli, and a stimulus set containing only 16 faces elicited significant activation in 80% of all cells. The cells in this patch may be performing the "structural encoding stage" of face processing (27); at this stage, faces are analyzed in terms of structural properties and semantic identity has not yet been made explicit.

Several previous single-unit studies have described a scattered clustering of face-selective cells in the temporal lobe (8,17,28), suggesting an underlying architecture of clumps (29) or columns (28), although these clusters may be larger (0.5 to 2 mm) than classical columns found in early sensory areas (7,30). Furthermore, optical imaging studies have found ~1-mm spots in anterior inferotemporal cortex that are selective for faces (31). Large parts of the temporal lobe may indeed be tessellated by columns selective for different kinds of objects. However, because of its reproducible anterior-posterior location and selectivity properties (both single-unit and LFP) across animals and its relatively large size (~16 mm^2), the middle face patch appears to constitute a different level of functional organization: a discrete area dedicated to face processing.

Why is it important that the brain contains an area consisting entirely of face-selective cells? First, this indicates that the brain uses a specialized region to process faces. Second, no brain region has previously been identified that is selective for a single visual form; in this sense, the fMRI face patches are analogous to the widely studied areaMT/V5, which is specialized for processing visual motion. Third, the finding that essentially all cells within this region were face-selective implies that either all the inputs are already face-selective, or a face-selective output can be generated from non– or partially face-selective inputs in just one step. Fourth, the fact that fMRI and single units were both specific for the same visual features confirms and extends previous evidence that the hemo-dynamic signal of fMRI can be highly correlated with single-unit activity, in higher order regions (32) as well as in lower tier regions (33). Fifth, the fact that many face-selective cells also showed a weak response to round clocks and fruits indicates that domain-specific face processing emerges at an early stage in form processing. And lastly, the grouping together of so many face-selective cells reiterates the advantages of modular architecture: An area consisting entirely of face-selective cells could achieve the richness of interconnections between large numbers of face-selective cells necessary to support holistic face processing.

References and Notes

1. Gross, CG. Handbook of Sensory Physiology. Jung, R., editor. Vol. vol. VII. Berlin: Springer-Verlag; 1973. p. 451

2. Yovel G, Kanwisher N. Neuron 2004;44:889. [PubMed: 15572118]

3. Haxby J, et al. Science 2001;293:2425. [PubMed: 11577229]

4. Gauthier I, Skudlarski P, Gore JC, Anderson AW. Nat. Neurosci 2000;3:191. [PubMed: 10649576]

5. Kanwisher N, McDermott J, Chun M. J. Neurosci 1997;17:4302. [PubMed: 9151747]

6. Perrett DI, Rolls ET, Caan W. Exp. Brain Res 1980;47:329. [PubMed: 7128705]

7. Perrett DI, et al. Hum. Neurobiol 1984;3:197. [PubMed: 6526706]

8. Desimone R, Albright TD, Gross CG, Bruce C. J. Neurosci 1984;4:2051. [PubMed: 6470767]

9. Baylis G, Rolls E, Leonard C. J. Neurosci 1987;7:330. [PubMed: 3819816]

10. Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB. Nat. Neurosci 2003;6:989. [PubMed: 12925854]

11. Pinsk MA, Desimone K, Moore T, Gross CG, Kastner S. Proc. Natl. Acad. Sci. U.S.A 2005;102:6996. [PubMed: 15860578]

12. To compute the size of the middle fMRI face patch in each monkey, we counted, for each hemisphere, the number of face-selective voxels parallel to the cortical surface in each slice intersecting the middle fMRI face patch (defined as the contiguous face-selective region in the STS at A6). We then multiplied the resulting voxel count by the voxel area (1.25 × 1.25 mm^2). Data were averaged over five monkeys.

13. Vanduffel W, et al. Neuron 2001;32:565. [PubMed: 11719199]

14. A significant response was defined as a reproducible response greater than baseline by at least 1 SD of the baseline firing for 20 consecutive bins.

15. Figure S5 shows the response profiles of all visually responsive cells in each monkey, sorted by the grid hole from which the cells were recorded (three grid holes in monkey M1 and three grid holes in monkey M2). In monkey M1, all three grid holes were loaded with face-selective cells. Similarly, in monkey M2, in both of the densely sampled grid holes, almost all visually responsive cells were face selective. Thus we can confidently conclude that the center of the fMRI face patch consists almost entirely of face-selective cells, and moreover this central region is not confined to a single millimeter-sized column, but covers a region at least 2 to 3 mm wide.

16. Baylis GC, Rolls ET, Leonard CM. Brain Res 1985;342:91. [PubMed: 4041820]

17. Perrett DI, et al. Proc. R. Soc. London Ser. B 1985;223:293. [PubMed: 2858100]

18. Tsao DY, et al. data not shown.

19. Logothetis NK. Curr. Opin. Neurobiol 2003;13:630. [PubMed: 14630229]

20. Rosch E, Mervis C, Gray MD, Johnson DM, Boyes-Braem P. Cogn. Psychol 1978;8

21. Sugase Y, Yamane S, Ueno S, Kawano K. Nature 1999;400:869. [PubMed: 10476965]

22. Liu J, Harris A, Kanwisher N. Nat. Neurosci 2002;5:910. [PubMed: 12195430]

23. Young MP, Yamane S. Science 1992;256:1327. [PubMed: 1598577]

24. Rolls ET. Philos. Trans. R. Soc. London Ser. B 1992;335:11. [PubMed: 1348130]

25. Eifuku S, De Souza WC, Tamura R, Nishijo H, Ono T. J. Neurophysiol 2004;91:358. [PubMed: 14715721]

26. Materials, methods, and supporting text are available as supporting material on *Science* Online.

27. Bruce V, Young A. Br. J. Psychol 1986;77:305. [PubMed: 3756376]

28. Tanaka K, Saito H, Fukada Y, Moriya M. J. Neurophysiol 1991;66:170. [PubMed: 1919665]

29. Perrett DI, Hietanen JK, Oram MW, Benson PJ. Philos. Trans. R. Soc. London Ser. B 1992;335:23. [PubMed: 1348133]

30. Harries M, Perrett DI. J. Cogn. Neurosci 1991;3:9.

31. Wang G, Tanifuji M, Tanaka K. Neurosci. Res 1998;32:33. [PubMed: 9831250]

32. Mukamel R, et al. Science 2005;309:951. [PubMed: 16081741]

33. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Nature 2001;412:150. [PubMed: 11449264]

34. We thank B. Moghaddam and T. Poggio for discussions; D. Hubel, C. Gross, and G. Yovel for comments on the manuscript; T. Chuprina, D. Freeman, S. Moeller, N. Nallasamy, N. Schweers, W. Vanduffel, and members of the Massachusetts General Hospital monkey fMRI group for technical assistance; and A. Dale and A. van der Kouwe for allowing us to use their multiecho sequence and undistortion algorithm. This work was supported by NIH (grant nos. EY13135 and EY13025), P41:RR14075, and the Mental Illness and Neuroscience Discovery (MIND) Institute. D.Y.T is supported by a Sofia Kovalevksaya Award from the Alexander von Humboldt Foundation.
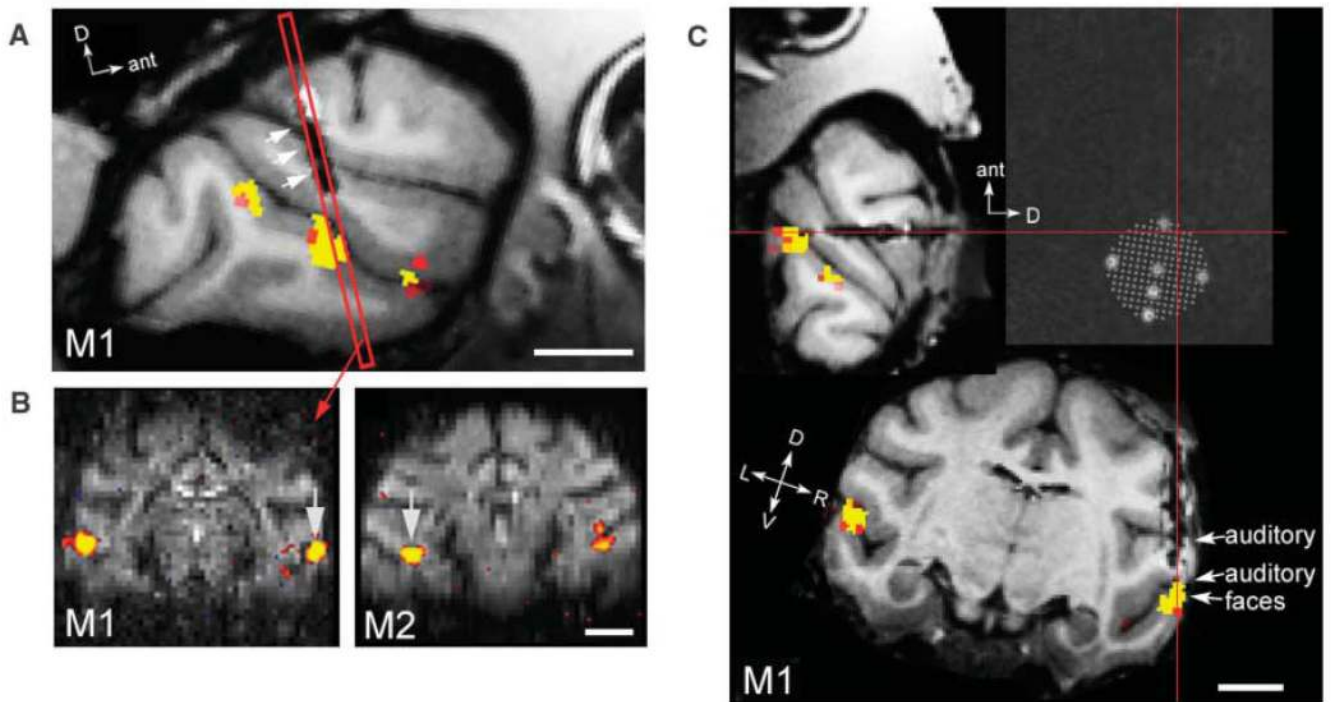
Fig. 1. Targeting an fMRI-identified face patch for single-unit recording
(**A**) A semi-sagittal section through the right hemisphere of monkey M1 showing three face-selective patches along the STS. Single-unit recordings in monkeys M1 and M2 were targeted to the middle face patch, located ~6 mm anterior to the interaural line; the red rectangle indicates a coronal slice passing through the middle face patch. The three white arrows point to the lesion left by the recording guide tube. (**B**) Two coronal slices showing the middle face patch in monkeys M1 (left) and M2 (right) (at A6.5 and A5.5, respectively). MION activation is overlaid on raw functional echo planar (EPI) images. Arrows point to the specific region targeted for electrophysiology in each monkey. In monkey M1, the targeted face patch was located on the lower lip of the STS in the right hemisphere. In monkey M2, the targeted face patch was located in the fundus of the STS in the left hemisphere. (**C**) Face patches and guide-tube track in three dimensions, rotated into the coordinate system of the recording grid (monkey M1). After chamber implantation, a high-resolution anatomical scan was obtained with six oil-filled markers positioned inside a grid in the recording chamber. We determined which grid hole to use by rotating the brain, together with registered face-selective fMRI activation, into the coordinate system defined by these markers. This panel shows three orthogonal slices passing through the point marked by the intersection of the red lines. Cells in this monkey were recorded from the hole at the intersection of the red lines, and from two adjacent, more-medial holes in the same row of the grid. The dark elongated lesion confirms that the guide tube passed through this point to accurately and precisely target the middle face patch. We recorded from all cells encountered between the start of the gray matter and the start of the white matter in the lower bank of the STS. Scale bar, 1 cm. $P < 10^{-4}$ for MION activations.
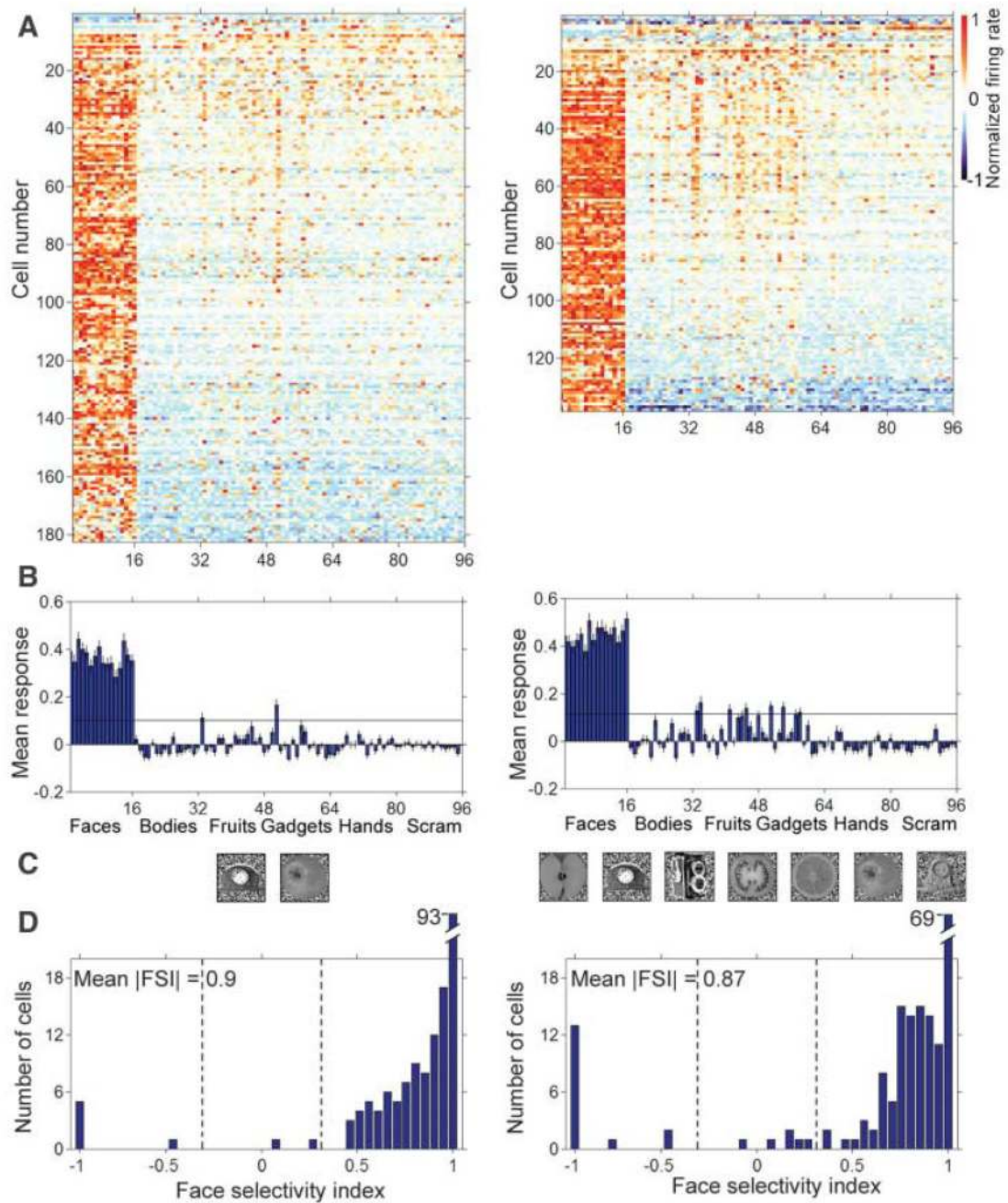
Fig. 2. Face selectivity of single units in the middle face patch

(**A**) Selectivity profiles of all visually responsive cells in monkeys M1 (left) (182 cells) and M2 (right) (138 cells) to 96 images of faces, bodies, fruits, gadgets, hands, and scrambled patterns (16 images per category, see fig. S1 for stimuli). Each row represents one cell and each column one image. The rows were sorted by the FSI, and the columns were sorted by image category. To compute selectivity profiles for each cell, responses to the 96 images were averaged over a 200-ms interval starting at the response latency, the baseline (the average response from 0 to 50 ms) was subtracted, and the response normalized. The average response time course to each of the 96 images is shown in fig. S6. (**B**) Average response to each of the 96 images across all visually responsive cells in monkeys M1 and M2. Error bars represent ±1

SE. The black line indicates six average SEs. (**C**) Nonface images that elicited a response above six average SEs in monkeys M1 and M2. Images are sorted from left to right by decreasing elicited response magnitude. (**D**) Distribution of FSIs across all visually responsive cells. Dotted lines indicate |FSI| = 0.33 (corresponding to a 2:1 ratio of face-to-nonface object response).
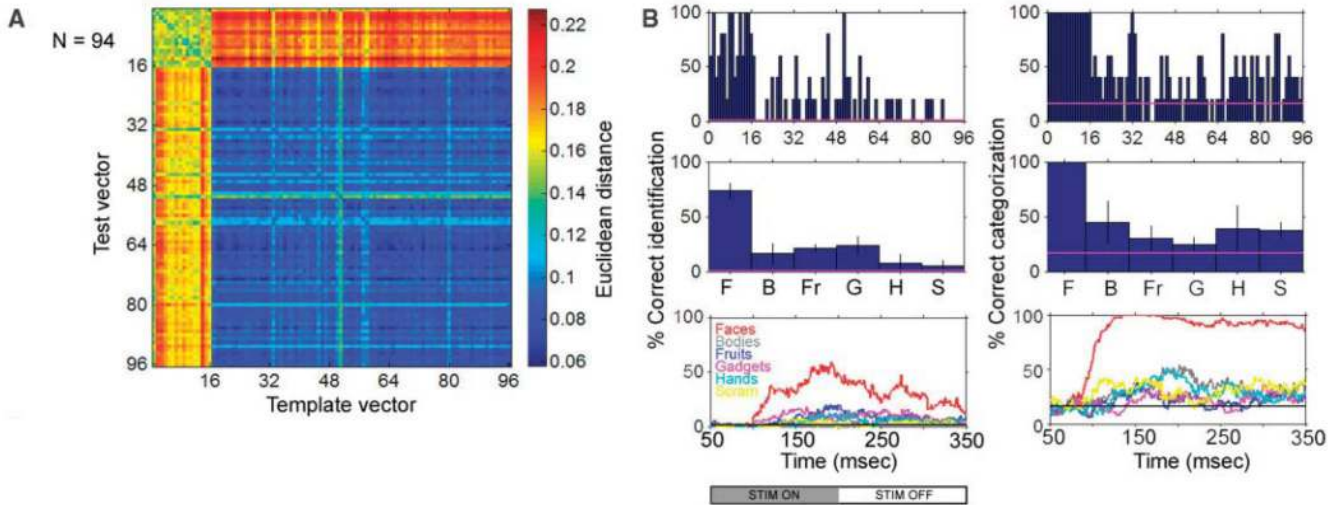
Fig. 3. Face identity and category information are carried by the population of face-selective cells
(**A**) Matrix of Euclidean distances between each of the 96 × 96 pairs of test and template response vectors. The maximum possible distance between a test and a template vector is $\sqrt{94}$. (**B**) The top row shows the percent correct identification and categorization for each image, based on a nearest neighbor algorithm. The middle row shows the same data grouped by category (F, faces; B, bodies; Fr, fruits; G, gadgets; H, hands; S, scrambled patterns). Error bars represent ±1 SE. The bottom graphs are the percentages of correct identification and categorization for six different categories as a function of time after stimulus presentation, computed using a 50-ms sliding window. Chance performance would be 1/96 for identification and 1/6 for categorization (indicated by the horizontal line in each graph). Data are from monkey M1.