

Single-Unit Activity during Natural Vision: Diversity, Consistency, and Spatial Sensitivity among AF Face Patch Neurons

David B.T. McMahon,^{1,2} Brian E. Russ,¹ Heba D. Elnaïem,¹ Anastasia I. Kurnikova,¹ and  David A. Leopold^{1,3}

¹Laboratory of Neuropsychology, National Institute of Mental Health, NIH, Bethesda, Maryland 20982-4435, ²Laboratory of Sensorimotor Research, National Eye Institute, NIH, Bethesda, Maryland 20982-4435, and ³Neurophysiology Imaging Facility, National Institute of Mental Health, National Eye Institute, and National Institute of Neurological Disorders and Stroke, NIH, Bethesda, Maryland 20982-4435

Several visual areas within the STS of the macaque brain respond strongly to faces and other biological stimuli. Determining the principles that govern neural responses in this region has proven challenging, due in part to the inherently complex stimulus domain of dynamic biological stimuli that are not captured by an easily parameterized stimulus set. Here we investigated neural responses in one fMRI-defined face patch in the anterior fundus (AF) of the STS while macaques freely view complex videos rich with natural social content. Longitudinal single-unit recordings allowed for the accumulation of each neuron's responses to repeated video presentations across sessions. We found that individual neurons, while diverse in their response patterns, were consistently and deterministically driven by the video content. We used principal component analysis to compute a family of eigenneurons, which summarized 24% of the shared population activity in the first two components. We found that the most prominent component of AF activity reflected an interaction between visible body region and scene layout. Close-up shots of faces elicited the strongest neural responses, whereas far away shots of faces or close-up shots of hindquarters elicited weak or inhibitory responses. Sensitivity to the apparent proximity of faces was also observed in gamma band local field potential. This category-selective sensitivity to spatial scale, together with the known exchange of anatomical projections of this area with regions involved in visuospatial analysis, suggests that the AF face patch may be specialized in aspects of face perception that pertain to the layout of a social scene.

Key words: eigenneurons; fMRI; macaque; movies; natural vision; single units

Introduction

Primate social exchange depends upon the real-time visual analysis of faces and bodies, their actions, spatial relationships, and social significance (Barraclough and Perrett, 2011). fMRI mapping in macaques has revealed that several regions within the inferotemporal cortex respond more strongly to faces than to other structured stimuli (Tsao et al., 2003; Pinsk et al., 2005; Bell et al., 2009; Ku et al., 2011). In turn, the majority of neurons in these fMRI-defined face patches respond categorically to images of faces (Tsao et al., 2006; Bell et al., 2011; Issa et al., 2013; McMahon et al., 2014b). More specific aspects of neural response selectivity vary between the patches, for example, in their prefer-

ences for specific features or tolerance to different viewing angles (Freiwald et al., 2009; Freiwald and Tsao, 2010; Issa and DiCarlo, 2012). Macaque fMRI studies using video presentation have highlighted the strong contribution of movement in the face patches (Furl et al., 2012; Polosecki et al., 2013; Russ and Leopold, 2015). One recent fMRI study used free viewing of social videos to obtain maps of face-selective cortex that closely matched those observed with block-design testing (Russ and Leopold, 2015).

Natural stimulus paradigms offer a complementary, data-driven approach to conventional stimulus presentation, as they permit one to investigate the relative neural contribution of different stimulus factors that are normally superimposed. Here we apply this approach to investigate single-unit activity in the anterior fundus (AF) face patch of the STS in monkeys viewing natural videos. We addressed three basic questions. First, to what extent does the repeated, free viewing of the same natural video elicit the same neural activity pattern across trials? Second, to what extent is the activity evoked from a given video shared among a group of tightly (<1 mm³) clustered AF neurons? Third, what specific stimulus features determine the activity of AF neurons during viewing of natural movies? In three monkeys, we recorded single-unit responses to multiple, repeated presentations of three 5 min videos, in each case exploiting the ability to maintain the isolation of single neurons over multiple sessions (Bondar et al., 2009; McMahon et al., 2014a, b).

Received Sept. 12, 2014; revised Jan. 13, 2015; accepted Feb. 15, 2015.

Author contributions: D.B.T.M. and B.E.R. designed research; D.B.T.M., H.D.E., and A.I.K. performed research; D.B.T.M., B.E.R., H.D.E., A.I.K., and D.A.L. analyzed data; D.B.T.M. and D.A.L. wrote the paper.

This work was supported by the Intramural Research Programs of the National Institute of Mental Health, National Institute of Neurological Disorders and Stroke, and National Eye Institute and by National Institutes of Health Grant EY018028 (to D.B.T.M.). We thank Kadharbatcha S. Saleem for helpful discussions and Hannah Johnson for assistance with data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to David B.T. McMahon, Laboratory of Sensorimotor Research, National Eye Institute, Building 49, Room 2A5049, 49 Convent Drive, Bethesda MD 20982hyphen4435. E-mail: mcmahond@mail.nih.gov.

DOI:10.1523/JNEUROSCI.3825-14.2015
0270-6474/15/355537-12\$15.00/0

Individual neurons in AF were highly reliable in their activity time courses across viewings, even when repetitions were separated by several days. In contrast, neighboring neurons were surprisingly uncorrelated, with only a minority showing a high degree of response correlation. We found that the spatial layout of animals depicted in the scenes, more than anything else, predicted the neural response time course. We discuss these findings in terms of the possible role of AF in integrating social and spatial stimulus information, as well as its position within the larger network of macaque face selective areas.

Materials and Methods

Subjects

Three rhesus macaques were implanted with chronic microwire electrode bundles held within an MRI-compatible chamber and microdrive assembly. The apparatus, surgical implantation protocol, and procedure for advancing the electrodes post surgically to search for suitable deep recording structures have been described previously (McMahon et al., 2014a). Laboratory designations of the subjects were SI (monkey 1, female, 5.0 kg), RH (monkey 2, male, 8.7 kg), and TO (monkey 3, female, 5.6 kg). Monkey 3 received bilateral electrode implants, whereas monkeys 1 and 2 received left and right hemisphere implants, respectively. All procedures were approved by the Animal Care and Use Committee and complied with the regulations of the NIMH and NIH.

Functional imaging

Functional and anatomical MRI scanning was carried out in the Neurophysiology Imaging Facility Core (NIMH, NINDS, NEI). Temporal lobe face patches were localized in all three monkeys using a standard fMRI block design contrasting epochs of face and nonface objects conducted using a 4.7 tesla vertical scanner (Bruker BioSpin). The details of the stimulus presentation, functional image acquisition, and data analysis have been described previously (McMahon et al., 2014b). Briefly, each functional scan consisted of 12 blocks of 24 s duration in which either a series of faces, body parts, inanimate objects, or no stimuli were presented. The last of these served as a fixation baseline condition. The stimulus set was the same as that used in a previous study for the same purpose (Bell et al., 2009). In all blocks the animal was rewarded with juice or water every 2 s for maintaining fixation within a 3 degree window around the fixation point. Hemodynamic responses were enhanced using the contrast agent monocrySTALLINE iron oxide nanoparticles injected intravenously before commencing data collection (Leite et al., 2002). Data were collected using either a two-channel surface coil (monkeys 1 and 3) or an 8-channel volume coil (monkey 2) and analyzed using AFNI (Cox, 1996) and custom software written in MATLAB (MathWorks).

Experimental design

Visual stimuli consisted of three movies depicting macaque monkeys engaged in natural behavior and interacting with humans. The content of the movies was assembled from commercially available documentaries and wildlife footage in a manner described in detail previously (Russ and Leopold, 2015). The movies were presented within a rectangular frame of 10.4° wide and 7.6° high (640 × 480 pixel resolution). Head position was stabilized throughout the experimental session by means of a chronically implanted head post. The monkeys sat in a primate chair in front of an LCD monitor and watched the movies while eye-position signals were recorded using an infrared video-tracking system (EyeLink; SR Research). Two monkeys (monkeys 1 and 2) received a water or juice reward for maintaining fixation within a large fixation window of 10.5°, which included the entire movie frame. In a series of recordings in monkey 3, static images were presented while the animal maintained fixation. Each stimulus appeared at the fovea for 300 ms, followed by a 300 ms interstimulus interval. Images were drawn from a large library comprising 1000 exemplars from 10 stimulus categories as described previously (McMahon et al., 2014a).

Physiological recordings

Recordings were obtained from bundles of 32 NiCr or 64 NiCr microwires chronically implanted within the anterior superior temporal

sulcus aimed at the AF face patch localized through fMRI scans (McMahon et al., 2014b). The microwire electrodes were designed and initially constructed by Dr. Igor Bondar (Institute of Higher Nervous Activity and Neurophysiology, Moscow, Russia) and subsequently manufactured commercially (Microprobe). Physiological recordings were performed in a radio frequency-shielded room. Data were collected using either a Multichannel Acquisition Processor (Plexon) with 32-channel capacity or a RS4 BioAmp Processor (Tucker-Davis Technologies) with 128-channel capacity. A gold wire inserted into a skull screw was used for ground and reference. Broadband electrophysiological responses were collected, from which individual spikes were extracted and analyzed after filtering off-line between 300 and 5000 Hz. Local field potentials (LFPs) were extracted off-line from the same broadband signal by low-pass filtering < 300 Hz. Off-line digital filtering of both the spike-band and LFP-band signals was done using a second-order Butterworth filter and the MATLAB function `filtfilt.m`. The gamma component of the LFP signal was extracted by filtering between 40 and 60 Hz and then rectifying each band-limited single-trial response. After determining that the low-frequency signals recorded on different wires were largely redundant (median cross-channel correlation coefficient, $r = 0.97$), we took the traces from a single channel as the representative LFP. All aspects of the task related to timing of stimulus presentation, eye-position monitoring, and reward delivery were controlled by custom software courtesy of David Sheinberg (Brown University, Providence, RI) running on a QNX computer. Movies were displayed using the Psychophysics Toolbox (Brainard, 1997) running on a separate PC.

Longitudinal tracking of single units

The method for spike sorting and longitudinal identification of neurons across days is described in detail previously (Bondar et al., 2009; McMahon et al., 2014a). Briefly, spikes from single units were sorted using the OfflineSorter software package (Plexon) by projecting waveforms into principal component space and identifying isolated clusters. Spikes recorded from the same channel on different days routinely had closely matching waveforms and interspike interval histograms, and were provisionally inferred to arise from the same neurons across days. This initial classification based purely on waveform features and spike statistics was tested against the pattern of stimulus selectivity and temporal structure of the neurons' firing evoked by visual stimulation. Guided by our previous observations that neurons in inferotemporal (IT) cortex respond consistently to statically presented visual stimuli across days and even months (Bondar et al., 2009; McMahon et al., 2014a), we used the distinctive visual response pattern generated by isolated spikes as a neural "fingerprint" to further disambiguate the identity of single units over time.

Data analysis

Behavioral and neural correlations. Analysis of oculomotor behavior and neuronal responses was conducted using programs custom written in MATLAB. For behavioral measures, periods of stability in horizontal and vertical eye-position traces were identified by detecting saccades that exceeded a velocity threshold of 20 degrees/s. The 5 min record of eye-position traces obtained upon each presentation of a movie was divided into epochs of viewing and nonviewing based on whether or not the eyes fell within the movie display rectangle. Correlations in the temporal structure of eye positions measured on different trials were computed as Pearson's coefficients after masking out the time periods corresponding to saccades and nonviewing epochs.

For neuronal analysis, LFP gamma activity time courses and spike firing rate histograms were displayed for each 5 min movie after convolving the average signal over trials from the same movie with a 500 ms Gaussian kernel. Correlations between time series of neuronal signals were computed as Pearson correlation coefficients. To assess the degree of correlation of single-neuron responses to movies across different runs, and contrast with the correlation in response structure across neurons, the data were divided into split halves comprising odd and even trials.

Construction of eigenneurons. The dimensionality of the dataset of neuronal responses, in other words the number of single units, was reduced by means of principal components analysis applied first to a matrix of N

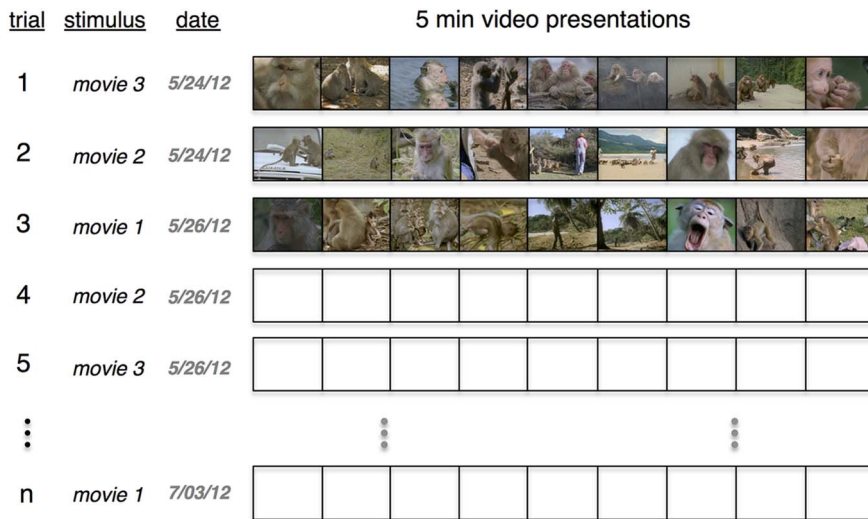


Figure 1. Stimulus presentation paradigm. Three 5 min movies, consisting of a range of scenes extracted from nature movies, were presented to each of the three animals repeatedly. A typical recording session, in which other experiments were also conducted, involved 12–16 movie presentations.

neurons by 90,000 time points (900 s of movie data divided into 10 ms bins). This computation yielded two measurements of interest. First, a matrix of $N \times 90,000$ new time series was generated through orthogonal rotation of the original response matrix. These dimensionally transformed neuronal responses constituted the eigenvectors, or “eigenneurons,” of the new space that could be used to reconstruct the original data by linear combination. These eigenneurons were ranked according to the proportion of overall shared variance that they explained. Second, the analysis yielded a vector of N scores for each of N neurons that indexed the extent to which each eigenneuron time course contributed to the original activity time course of each neuron. The eigenneuron time courses were used to visualize the response patterns inherent in the population of neurons in a dimensionally reduced format. The first two dimensions of the N scores were further used to visualize the degree of clustering among neurons in principal component space.

Relationship of neural activity to visual content. The relation between stimulus content and neuronal responses was examined visually by constructing a library of display movies. The movies used for the purpose of data visualization were constructed by integrating the footage initially viewed by the animals with scrolling raster plots and spike time sound tracks of individual neurons. Data display movies were assembled frame by frame in MATLAB and compiled using QuickTime Pro 7.6.6 software (Apple).

Specific hypotheses regarding the influence of movie content on spikes were modeled as a matrix of movie content parameters describing the movie content in each of 175 scenes. For this analysis, a single scene was defined as the time interval between cuts to different camera shots, which lasted 5.1 s on average (min 0.4 s, max 24.2 s). The middle-most frame of each scene was selected for measurement, and the following features, if present on the screen, were outlined manually using a custom-written MATLAB program: heads (almost always faces), forelimbs (arms and hands), torso, and hindquarters (including tail, buttocks, and hind legs and feet). If more than one animal was depicted in a screen, the most prominent individual was selected for measurement. As in the eigenneuron calculation above, principal component analysis (PCA) was applied on the neural responses across the population, but in this case on a scene-by-scene basis. Specifically, these scene-based eigenvectors and their scores were computed based on a matrix of mean firing rate values for N neurons by 175 scenes. These eigenvectors could then be readily compared with the movie content, computed separately for each scene. A multiple factor regression analysis was conducted to assess the influence of the scene content regressors on scene-based eigenneurons. Significance of main effects for feature size and pairwise interaction effects was determined by the resulting t statistics for each factor.

Results

Three macaques repeatedly viewed a set of movies that contained a diverse range of footage culled from nature movies depicting macaques in their natural environments and interacting with humans. Two monkeys (monkeys 1 and 2) received periodic liquid reward simply for looking at the screen, and the third monkey (monkey 3) received no reward during the movie presentations. The animals were not otherwise required to perform any particular task, but were simply permitted to view the movies freely. Each subject was implanted with a bundle of microwire electrodes (McMahon et al., 2014a).

Each animal viewed each 5 min movie between 9 and 23 times (median for three monkeys: 22, 14, 10), with the movies presented in pseudorandom order (Fig. 1). Because of the stability of the recordings, we routinely collected trials while monitoring individual neurons on different

days, with one trial in this case constituting the presentation of a 5 min movie. Eye position was continuously tracked and recorded using a video-based system. The traces in Figure 2A show the scanning pattern of monkey 1 during the repeated viewing of a 1 min segment of footage. The individual horizontal and vertical eye-position traces are shown in gray, with the mean over eight viewings shown in black. To quantify the consistency across viewings, we adopted a correlational method from a previous study evaluating gaze similarity in humans and monkeys watching videos (Shepherd et al., 2010). In agreement with the results of that study, we found a positive correlation in gaze position for repeated viewings of the same movie, but no correlation across viewings of different movies (Fig. 2B). The within-movie correlations in eye traces were significant in each of the three animals ($r = 0.35$ in monkey 1, $r = 0.25$ in monkey 2, $r = 0.18$ in monkey 3, $p \ll 0.01$ in all cases, one-sampled t test) and in all cases substantially higher than the across-movie correlations (median $r = 0.001$, $N = 2337$ across-movie trials pairs). Eye-position signals were also significantly correlated between different monkeys watching the same movie (median $r = 0.13$, $N = 2004$ across-monkey trial pairs, $p \ll 0.01$, one-sampled t test). The animals' spontaneous behavior thus indicated that they were engaged by the movie content in a manner that was consistent across repeat viewings in the same animal and, to a lesser extent, consistent between animals. The following sections describe electrophysiological activity from a face-selective patch of cerebral cortex recorded while the animal was actively viewing the movies.

Responses to movies: consistent across days and diverse among neurons

For each monkey, electrode targeting was guided by presurgical fMRI scans that determined the location of AF relative to the network of temporal lobe face patches. Consistent with previous findings (Tsao et al., 2008; Pinsk et al., 2009) AF was readily identifiable within the anterior STS despite variation between monkeys, occupying in different animals either the fundus (monkey 1), the lower bank (monkey 2), or the upper bank of the STS. Along the rostrocaudal axis AF was localized at 16 mm (monkeys 1 and 3) or 19 mm (monkey 2) anterior to the interaural canal. Positioning of the MR-compatible electrode and microdrive in

the fMRI-defined face patch was verified by MRI post surgery (Fig. 3A). As the monkeys freely watched the movies, we recorded both single-neuron activity and LFP in area AF. The gamma-range (40–60 Hz) LFP activity was correlated across subjects watching the same movies (Fig. 3B; pairwise correlations between monkeys: $r_{m1m2} = 0.35$, $r_{m1m3} = 0.49$, $r_{m2m3} = 0.46$, all significant at $p \ll 0.01$, F tests). This result agrees with previously reported cross-subject coupling of LFP signals in the auditory cortex of human surgical patients presented with the same movie soundtrack (Mukamel et al., 2005).

The single-unit responses to individual movies were highly similar across viewings, and across sessions, as can be seen from three example neurons in Figure 3C. The consistency of these spiking responses across different sessions is similar to what we previously observed using statically presented images during passive fixation (Bondar et al., 2009; McMahon et al., 2014b). A conspicuous feature of the three simultaneously recorded neurons highlighted in Figure 3D is that their responses bear little similarity to one another. The peaks, troughs, and plateaus that are consistent across one neuron's responses do not appear to be shared with other neurons. The sharp contrast between the high within-neuron correlation across sessions and low between-neuron correlations can be seen in Figure 3E, where correlations were computed using split-halves analysis for the population of 69 neurons (16 in monkey 1, 5 in monkey 2, and 48 in monkey 3). Two independent mean time courses for each cell were computed by drawing odd and even trials of the same neuron, which on average were highly correlated [median $r = 0.73$, interquartile range (IQR) 0.52–0.86]. However, when the same analysis was applied to odd and even trials of different neurons, the correlation was much lower (median $r = 0.04$, IQR –0.04 to 0.15, $N = 2346$ pairs) but still significantly greater than zero ($p \ll 0.01$, one-sampled t test). Since the microwire tips within the bundle spread minimally (all tips falling within a 1 mm diameter; McMahon et al., 2014a), the heterogeneous response profiles evoked by the movies evidently reflect a high degree of diversity within a voxel-sized region of cortex. This observation fits with similar comparisons within and between sessions in AF and elsewhere in the inferotemporal cortex using static images (Bondar et al., 2009; McMahon et al., 2014b). In the next sections we investigate more specifically the types of visual content that drive the activity of single neurons in this area.

Categorizing neurons based on response similarity

What features of the movie content might account for neural activity in AF in a manner that is both reliable across viewings and very different among neighboring neurons? The longitu-

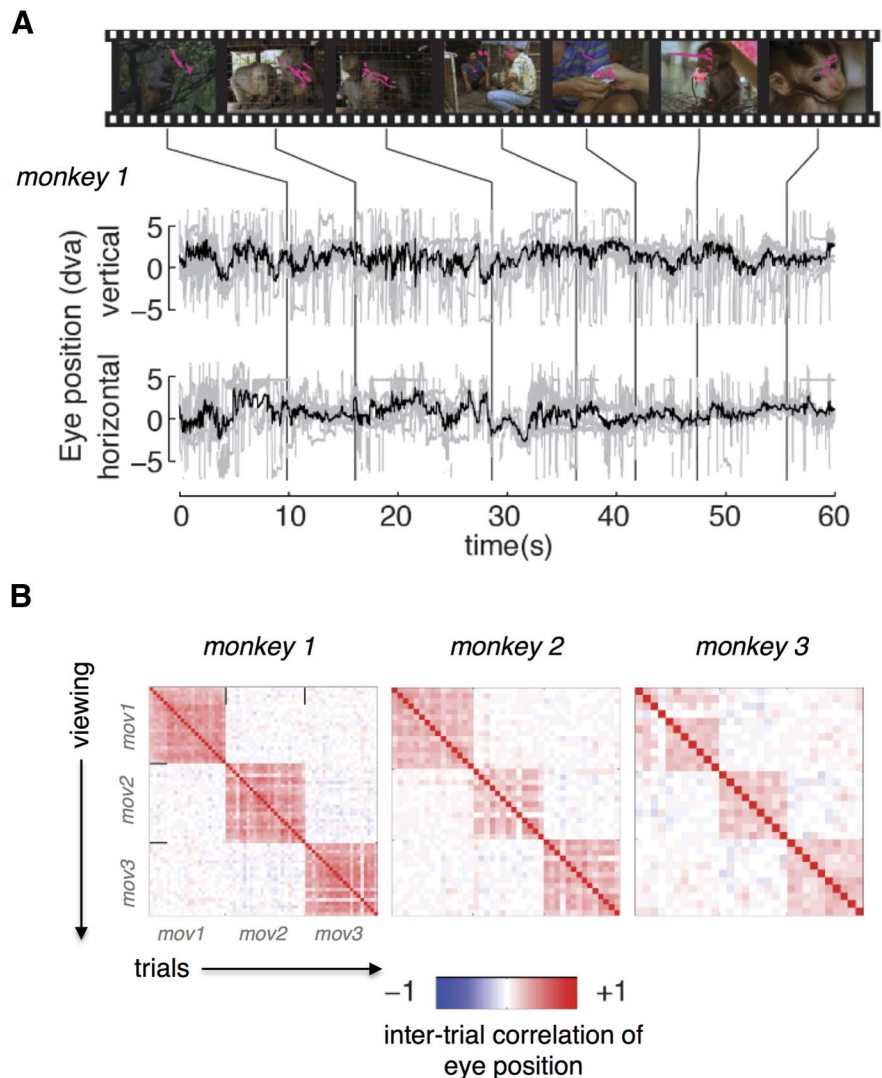


Figure 2. Consistency of gaze behavior during movie viewing. **A**, Eye position as a function of time recorded from monkey 1 during eight different viewings of Movie 1 (single trials shown by gray lines; mean in black). Top, Representative movie frames with overlaid x - y plot (magenta) showing eye-position signals recorded over the preceding 14 frames (468 ms). **B**, Correlation matrices assessing consistency in eye position across trials for each of the three monkeys. Each element in the matrix represents the correlation coefficient obtained for a pair of trials. Sorting trials by movie revealed that the eye-movement patterns within the movies were consistent (red squares along the diagonal) but that the eye-movement patterns across movies were not (white fields off the diagonal).

dinal recordings, and consequent accumulation of trials over multiple sessions, allowed us to gain good estimates of each neuron's mean responses to 15 min of movie scenes. The mean responses of the 69 neurons are displayed as a heat map in Figure 4A, with the ordering based on PCA, as described below.

We asked to what extent aspects of the movie-driven responses were shared across the neural population. Applying PCA to mean response time courses for each neuron generated a family of orthogonal basis functions, or eigenneurons, which best captured shared aspects of the spiking responses in the population. This analysis revealed that the overall dimensionality of the neuronal space in AF was relatively high (Fig. 4B), in that 26 PCs were required to account for 80% of the cumulative variance. Nonetheless, the first two PCs together accounted for ~24% of the overall variance (14 and 9%, respectively), indicating that neurons did share some aspects of their responses.

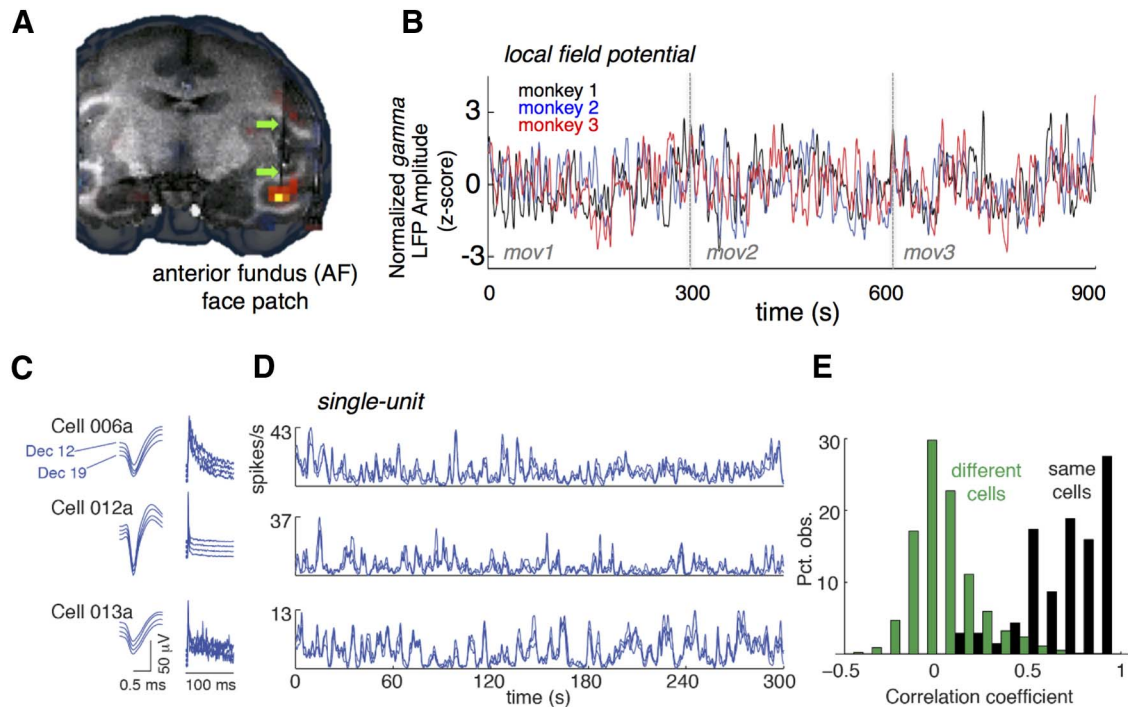


Figure 3. Neural responses in AF. **A**, Localization of the AF face patch in monkey 3 using fMRI. The shadow of the microwire electrode bundle can be seen inside the face-selective fMRI responses in AF. **B**, Mean gamma band of the LFP activity recorded from three monkeys across the three concatenated movies. The time course reveals that the LFP in the three monkeys' AF face patches is similarly modulated by the movies. **C**, Longitudinally recorded spike waveforms (left) and interspike interval histograms (right) for three neurons recorded over 8 d in monkey 2. **D**, Activity of the same three neurons in **C** over the entire run of Movie 1. Each line represents the mean firing rate obtained on a single day (3 or 4 trials per day). While the intersession modulation for each neuron is highly similar, the movie-driven time courses are very different in the three neurons. **E**, Distribution of split-halves correlation coefficients obtained between odd and even trials. Sample sizes: 69 neurons (same cells) and 2346 pairs (different cells).

The way in which this shared information was distributed can be seen in the ordering of the responses in Figure 4A. Here, the neural responses are ordered based on the decreasing value of the first PC (PC1) coefficient (eigenvalue). This ordering reveals that a number of neurons located near the top of the heat map (i.e., those with high PC1 coefficients) gave brief, coordinated responses at specific time points throughout the 15 min of total viewing time. These are apparent as brightly colored vertical stripes. A smaller number of neurons at the bottom of the heat map were likewise correlated in their activity, with more prolonged plateaus of coordinated activity that occurred less frequently. In the intermediate range, neural responses exhibited a broad range of response patterns that included some shared features, but generally appeared to be distinct. Thus informal inspection of this Figure 4A neural activity suggests there may be three groups of neurons that are relatively distinct in the way they respond to the movies. This grouping was also apparent when the eigenvalues for PC1 and PC2 were plotted as a scatterplot (Fig. 4C), prompting us to designate the neural groups in Figure 4A as clusters c1, c2, and c3.

To gain an additional perspective on shared activity across the population, we computed a matrix of pairwise correlation coefficients from the time courses of the ordered set of 69 neurons (Fig. 4D). The c1 neurons showed a relatively high level of correlation with one another (c1–c1 mean $r = 0.49$, IQR = 0.40–0.58), as did the c3 neurons to a lesser extent (c3–c3 mean $r = 0.27$, IQR = 0.10–0.33). In contrast, c2 neurons showed very low correlation (c2–c2 mean $r = 0.07$, IQR = -0.01 to 0.14) that differed significantly from those found in the other clusters ($p \ll 0.05$, two-sampled t test). Together, these results indicate that, despite an overall low correlation among neurons across the pop-

ulation, two groups (c1 and c3) stand out as sharing responses to key features of the videos.

Sensitivity to apparent viewing distance

What features give rise to the sharp, synchronized responses in the c1 cluster? Given that AF is defined by its face sensitivity, it is natural to expect that the appearance of faces are likely to be important. However, it seems unlikely that the explanation is the simple presence or absence of faces, since the movies are rich with faces throughout and very often multiple faces appear on the screen at the same time. The strong coordination of responses in the c1 cluster suggests that something more specific is responsible for the observed responses.

With the aim of understanding what aspects of the movie content gave rise to these distinct patterns of neuronal responses, we constructed display movies that combined the original footage viewed by the animals together with scrolling raster plots and spiking sound tracks of individual neurons. Movie 1 shows a sample of footage contrasting the responses of two particularly informative single units. The neurons were drawn from c1 (cell 102a) and c3 (cell 114a), and together illustrate the response opponency that characterized epochs of coordinated activity observed between the two clusters. In seeking to account for the alternating epochs of high and low spiking rates, the most obvious trend was the match between variation in neuronal responses and the viewing distance of animals depicted in the scene.

A portion of the neural data from the example movie described above is depicted in static form in Figure 5. This figure summarizes the first 22 scenes from Movie 1, comprising ~ 2 min. The spikes recorded over multiple viewings from each of the two neurons (Fig. 5A) illustrates the reliability of responses, along

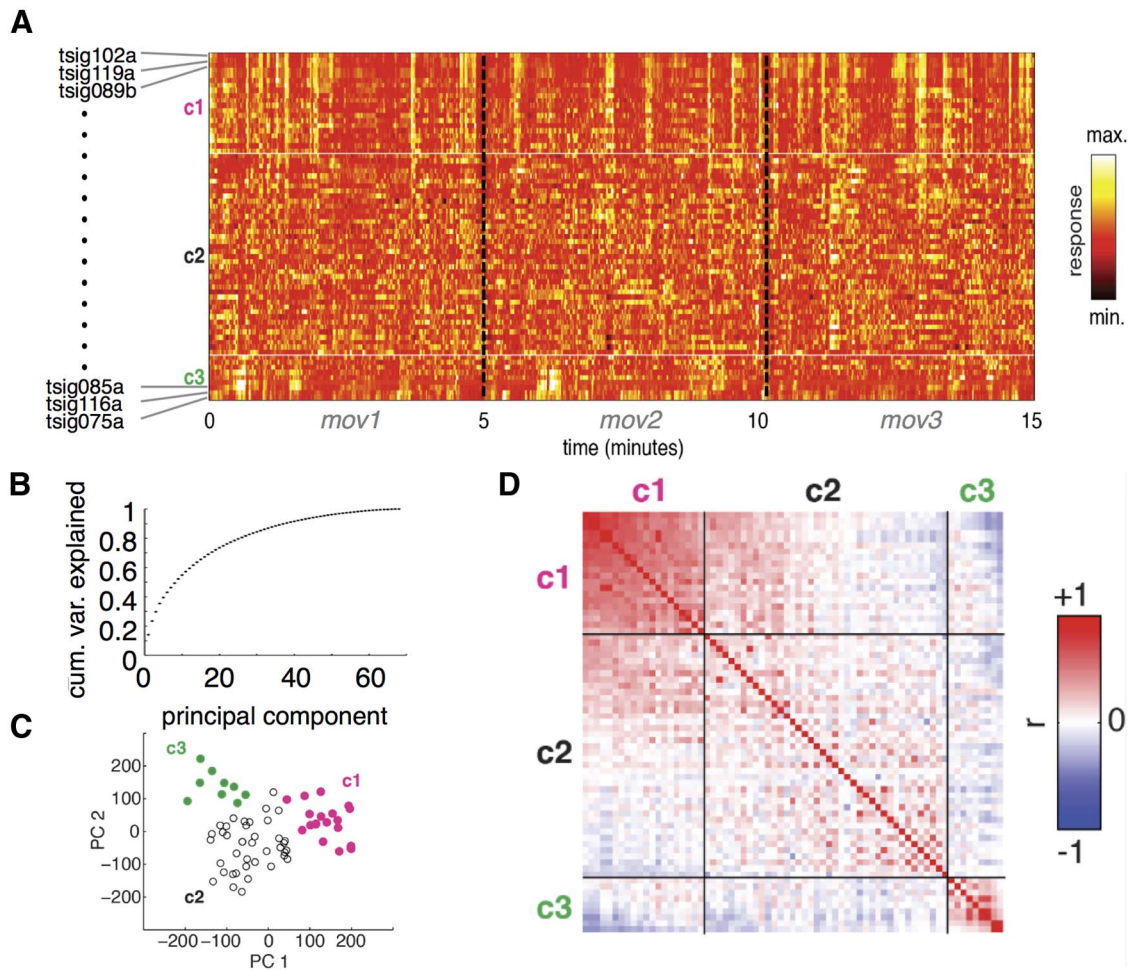
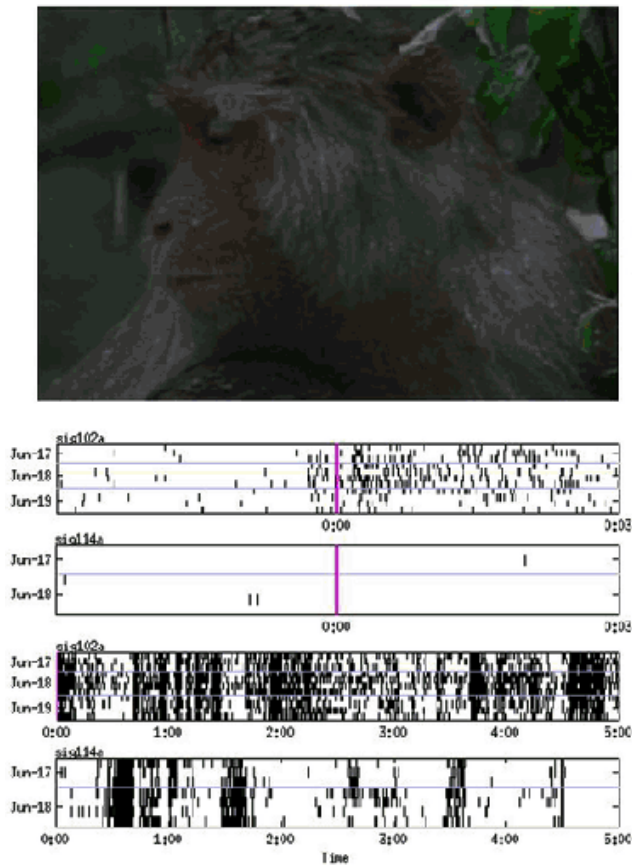


Figure 4. Structure in the responses of single units to movies revealed by principal components analysis. **A**, Heat map showing the mean responses of 69 different neurons to multiple presentations of all three 5 min movies. Each row corresponds to their decreasing PC1 eigenvalue. Horizontal lines indicate the boundaries between three groups of neurons (c1, c2, and c3) derived from the principal components analysis (see text and **C**). Vertical dashed lines indicate the points of concatenation of the three movies. **B**, Cumulative variance accounted for by all 69 principal components derived from the mean responses of 69 neurons analyzed in this study. **C**, Scatterplot of PC1 versus PC2 for all 69 neurons. Neurons showing similar response characteristics in **A** appear coarsely divided into clusters, designated as c1, c2, and c3. **D**, Correlation matrix of firing rate over time for 69 neurons, sorted first by membership in clusters shown in **C** and second by PC1 coefficient. Each element in the matrix represents the correlation coefficient in the mean firing rates of two neurons over the three movies concatenated together.

with what appear to be somewhat complementary firing across scenes, depicted by the colored background. The enumeration of individual scenes, based on camera cuts, is specified in greater detail in Figure 5*B*. Representative still images from each of the 22 scenes is shown in Figure 5*C*. Here the scenes are coarsely classified as near scenes, intermediate scenes, and far scenes based on the proximity of the monkeys or humans within the scene. The two neurons in this example were drawn from c1 (cell 102a) and c3 (cell 114a), and together illustrate the response opponency that characterized epochs of coordinated activity observed between the two clusters. While faces were clearly important features in determining the responses, close-ups of other body parts, such as hands, sometimes yielded equally strong responses (for example scene 7, starting at 0:40 in Movie 1). Similarly the time course of the first two eigen-neurons, EN1 and EN2, when compared with the scene categories, reveal that the shared activity in the population reflected whether animals depicted in the scene appeared to be near or far away (Fig. 5*D*). The highest values of EN1 coincided with close-ups of monkey faces, whereas those of EN2 were most commonly associated with distal views of animals, often whole bodies, but again typically including faces.

We thus formed the initial impression that many AF neurons were sensitive to the spatial layout of monkey and human faces within a scene. To more rigorously assess the video content underlying the responses of these neurons, we took the spatial scale of faces and other body parts as a proxy for apparent viewing distance. While viewing distance and spatial scale are separable variables, the real world absolute size of macaque faces has a limited range. Thus a macaque face subtending 12 visual degrees corresponds to an animal <1 m away, whereas a macaque face of 1.5 degrees corresponds to an animal nearly 4 m away. As the upper bank of the STS in the vicinity of AF is known to be interconnected with parietal and parahippocampal regions associated with spatial encoding (see Discussion), it is possible that neurons in this face patch are specialized to signal the spatial configuration of faces and other parts of the body within a visual scene. While neurons throughout the inferotemporal cortex are known to be sensitive to image scale, a particularly strong relationship between scale and neural responses in AF would stand as evidence in favor of this spatial encoding hypothesis.

To rigorously assess the relationship between neural firing and the scale of faces and other body parts in the videos, we traced



Movie 1. Responses of two neurons illustrating apparent near-far opponency. The top frame of this dynamic display plays 120 s of video content from one of the movies viewed by the subjects during the experiment. Scrolling raster plots show the responses of two neurons (sig102a and sig114a) at an expanded (top) and a compressed (bottom) timescale. Each row of the raster illustrates the neuron's activity during a single 5 min viewing of the movie. Trials recorded on different days are separated by horizontal lines. For the sake of clarity the same neuron responses to the same movie content are displayed in static format in Figure 5. The stereo sound track accompanying this display plays a spike on the left (sig102a) and right (sig114a) speakers whenever the corresponding neurons fired a spike on any trial during the experiment.

outlines around several features of interest found in all 175 scenes of the three movies (Fig. 6A; see Materials and Methods). From each scene we selected a representative still frame and computed the area of outlines drawn around the following features: faces, arms (including hands), torso, and hindquarters (including tail, rump, and hind legs). We likewise divided the neural data into scene epochs, computing the mean response to 175 scenes for each of the 69 neurons (Fig. 6B). Again applying the PCA, but now to this smaller scene-based matrix, we computed a family of scene-based eigenneurons (EN1_s, EN2_s, etc.) that, while similar to the time-based eigenneurons above, allowed for a more convenient comparison of population activity with aspects of scene content. This analysis revealed a striking correlation between the first eigenneuron EN1_s and the computed face area FA_s in each scene (Fig. 6C; $r = 0.78$, $p < 10^{-30}$, F test). Note that this analysis was not sensitive to the contrast between the presence and absence of a face, as scenes without faces were scored as dummy (NaN) variables. This finding, in which >60% of the response variance of the EN1_s signal can be explained by a single variable, indicates that spatial scale exerts a strong effect on this neural population. This result is consistent with the idea that neurons in face patch AF play an important role in representing the spatial positions of

socially relevant stimuli within a scene, as suggested by its anatomical connections.

Opponency between faces and other body parts

We examined the contribution of the other body features to neural responses to gain insight into whether faces alone were critical in shaping neural activity during the video scenes. Since we had extracted the spatial scale of arms, torso, and hindquarters, in addition to faces, we computed a multifactor regression analysis modeling the effects of the spatial scale of all four body features on neuronal activity. As expected based on the analysis of faces alone, we found that the scale of the body features was closely linked to the variation in EN1_s and EN2_s, with the full regression model accounting for 64 and 43% of their total variance, respectively. Further examination of individual features and their interactions revealed that not only were faces special, exerting by far the strongest influence on neuronal activity (EN1_s $p = 10^{-19}$; EN2_s $p = 10^{-7}$, F test), but that other significant factors all showed negative coefficients, indicating that the neural modulation due to the scale of other body parts was opposed to that of faces (Fig. 7A, B). For EN1_s the only other significant effect was a negative correlation with the scale of animal hindquarters. For EN2_s, there was a negative relationship with the scale of torso, as well as with the interaction of both hindquarters and torso scale with face scale. These results suggest that, to the extent that the scale of other body parts influence neurons in AF, they do so by reducing what would otherwise be strong responses to faces.

As noted above, examination of the display movies conveys the intuitive impression that the neurons encode proximity of conspecifics in the scene. If this interpretation is correct, then the same neurons might also be expected to show sensitivity to other spatial relations in addition to viewing distance. We tested this possibility by computing a multiple regression analysis modeling the effect of face size and face location (left–right and up–down) relative to the movie frame. For EN2_s (but not EN1) there was a significant effect for the location of faces in the up–down direction only ($p = 0.01$). Thus in addition to their scale sensitivity, the neuronal population displayed at least some spatial sensitivity within the tangent plane of the visual field.

We further considered whether the effects we observed could be explained by spatial frequency tuning, rather than sensitivity to face size per se. To test this possibility, we computed the spatial frequency content of each of the representative frames drawn from the 175 movie scenes. The mean spatial frequency power spectrum (averaged vertically across pixels) was divided into quartiles corresponding to four spatial frequency bins: low (<7.75 cycles/deg), medium (upper limit 15.5 cycles/deg), high (upper limit 23.25 cycles/deg), and very high (>23.25 cycles per degree). We again computed a multiple regression analysis modeling face size and the mean spatial frequency within each of the four bins as factors. There was no significant relation between EN1 and any spatial frequency component ($p > 0.05$). The medium, high, and very high bins were all significantly related to EN2 at a significance level of 0.05. However, these effects were dwarfed by the magnitude of the effect of face size on EN2. We conclude that, although the AF population showed some sensitivity to spatial frequency, these relations were not of sufficient strength to account for the observed sensitivity to face scale.

Sensitivity of LFP signals to face and body part scale

We asked whether the influence of on-screen face scale on neuronal activity was also present in the LFP signals sampling local activity in the vicinity of AF. Indeed, we found a significant cor-

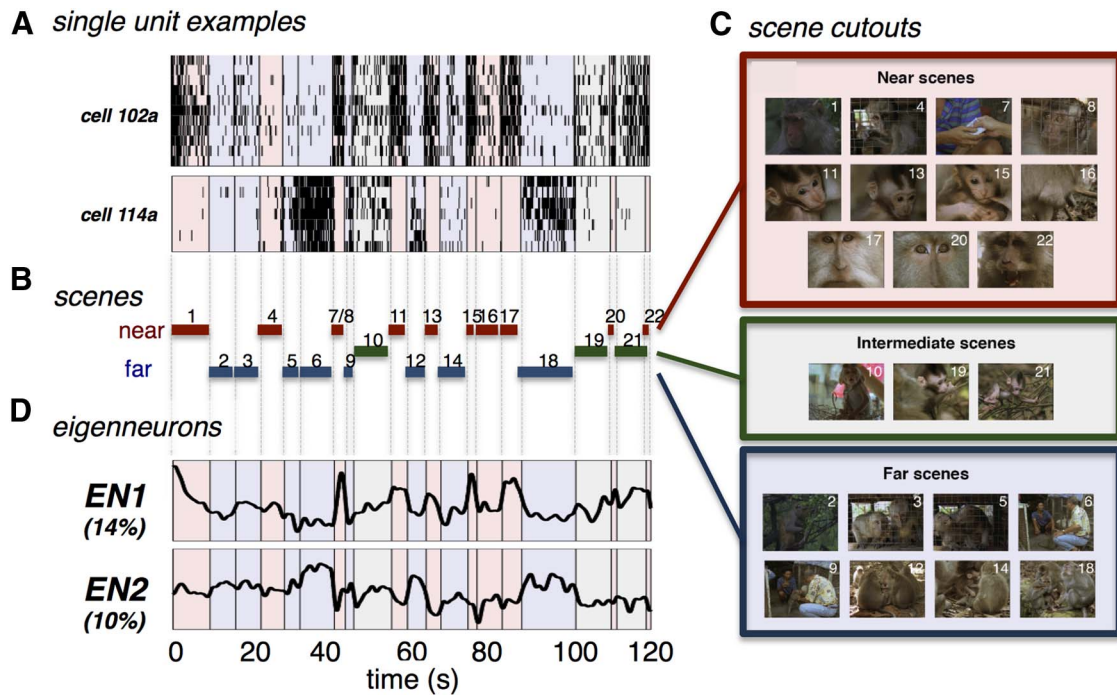


Figure 5. Scene-based comparison of spiking and movie content. **A**, Raster plots contrasting the spikes fired by two neurons (cell 102a from c1, and cell 114a from c3). Background colors correspond to those introduced in **C**. **B**, Classification of the 22 scenes from Movie 1 into near (red), intermediate (green), and far (blue) viewing distances. Numbers refer to the scenes whose still images are shown in **C**. **C**, The still images corresponding to the middle frame of each scene identified in **B**. **D**, Time course of first two eigenneurons (EN1, EN2) during the same scenes depicted above.

relation between face scale and the mean gamma amplitude observed over movie scenes (Fig. 7C; $r = 0.24$, $p = 10^{-8}$, F test). However, unlike the scene-based eigenneurons, the scene-based LFPs showed positive correlation with the scale of both the torso and hindquarters ($r = 0.21$ and 0.20 , respectively, $p < 10^{-3}$, F test). The basis of this difference is unknown, and may reflect the fact that the LFP is thought to be more sensitive to the summed synaptic potentials near the recording site than to the spiking output (Mitzdorf, 1987).

Assessment of scale and category selectivity of AF neurons for static images

The properties of AF neurons are relatively unexplored compared with other temporal lobe face patches. We therefore sought to clarify how the response patterns evoked by movies could be related to the category selectivity of AF neurons as determined by presentation of static images. We obtained responses from 30 visually driven neurons recorded in monkey 3, of which five neurons showed suppressive responses. Consistent with neurons from other face patches (Tsao et al., 2006; Issa and DiCarlo, 2012), the population-averaged activity of the 25 positively responding neurons was clearly the strongest for face stimuli (Fig. 8A). The same trend held true of almost all the individual neurons as well: all cells but one (96%) gave the strongest responses to either monkey faces (16 neurons) or human faces (8 neurons; Fig. 8B). After faces, the next strongest responses were evoked by whole monkey bodies, possibly because of the fact that these images typically included heads. We conclude that the degree of category selectivity of AF neurons is qualitatively comparable to the properties reported for other face patches (Tsao et al., 2006).

Previous recordings from regions of IT cortex that were not necessarily face selective revealed that neurons are commonly sensitive to image size (Op De Beeck and Vogels, 2000). In neurons showing both scale and stimulus selectivity both properties

are commonly independent, in that the rank ordering of a neuron's stimulus preferences is preserved across scale transformations (Ito et al., 1995). Broadly construed, the results of Ito et al. (1995) support the expectation that scale and stimulus preferences should sum linearly (their Fig. 9), as the stimulus preferences of IT neurons commonly do (Köteles et al., 2008; McMahon and Olson, 2009; Sripathi and Olson, 2010). Taken at face value, the eigenneuron sensitivities revealed by our scene-based multiple regression analysis (positively coupled to faces but negatively coupled to hind quarters and torso) appears to be at odds with the notion of linearly additive scale and object tuning. Upon further investigation with static images, the response patterns of AF neurons displayed category-dependent scale preferences (Fig. 8C). Many AF neurons showed a monotonically increasing relation between single-unit response and image size for monkey faces ($N = 21$) and a decreasing relation for all non-preferred image categories (butterflies, plants, man-made objects, scenes, and Fourier descriptor patterns; $N = 22$). A smaller subset of neurons ($N = 9$) displayed the inverse trend, namely responses to monkey faces that decreased monotonically with image size. Among the population of 30 neurons the slopes of the image scale gain factor were inversely correlated between monkey faces and nonfaces objects (linear regression, $R^2 = 0.27$, $p = 0.002$). Thus in contrast to IT cortex in general (where tuning for size and object preferences appear to be additive), neurons in the AF face patch show category-specific tuning for image size. When assessed with a two-way ANOVA modeling category and size as factors, all neurons but one (97%) had a significant interaction effect. Thus the effect of scale on AF neurons is strongly dependent on image category.

Discussion

Using a rich, relatively unconstrained regime of visual stimulation, we found that neurons in the AF face patch responded consistently

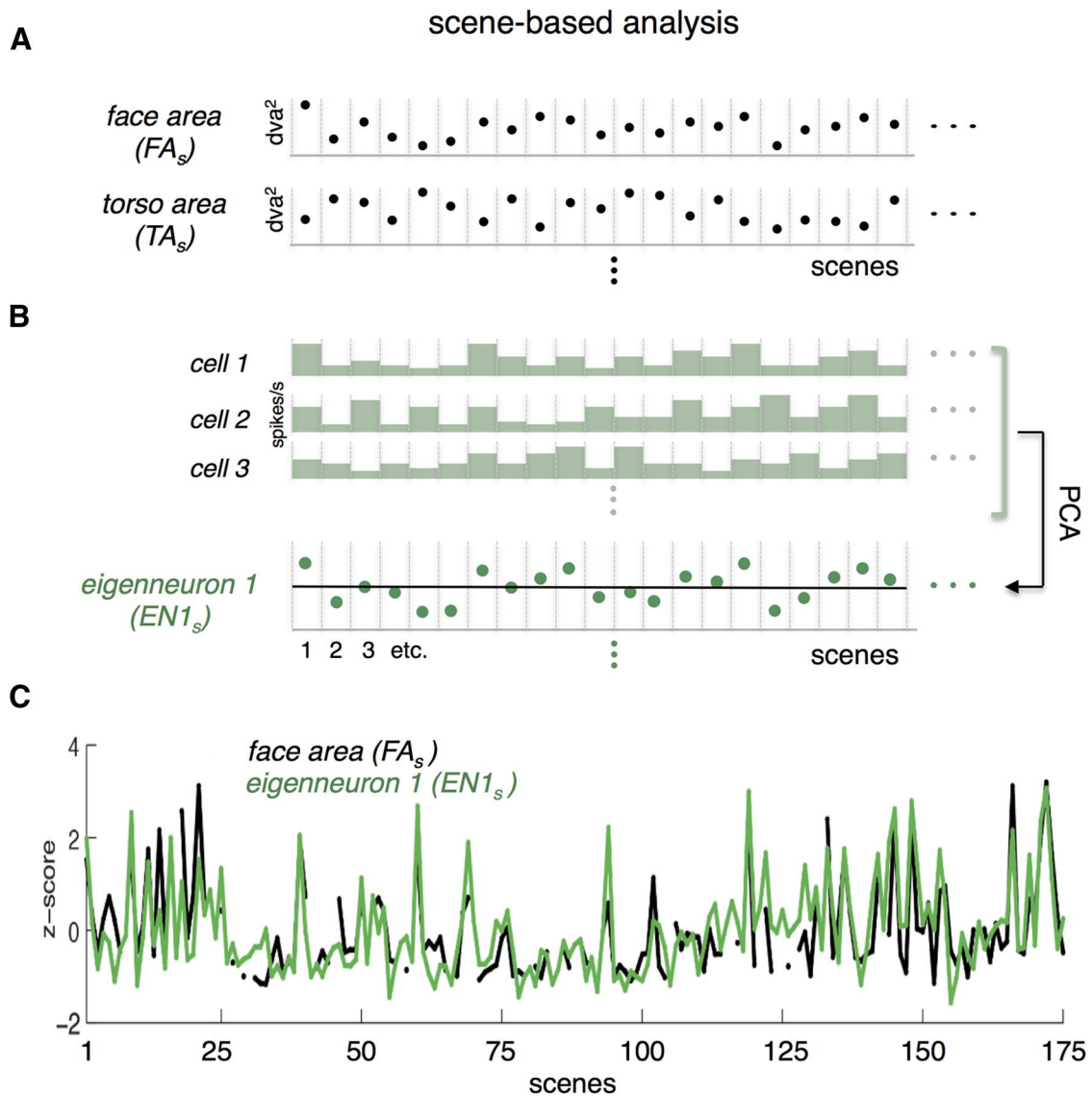


Figure 6. Detailed analysis of scene-based eigenneuron versus face area. **A**, For each scene in the three movies (175 scenes total) the mean area of the four body compartments (face, arm, torso, and hindquarters) was computed, forming a 175-element vector (see Materials and Methods). **B**, Spiking activity was similarly computed in a scene-based fashion for each cell, with the mean firing rate computed for each scene. PCA was then performed on the 69 arrays of 175-element spiking rate vectors. The resulting scene-based eigenvectors (e.g., $EN1_s$) could then be analyzed for correspondence with the scale of the body parts, and particularly the face. **C**, Strong correspondence between the mean face size for each scene and scene-based eigenneuron activity.

to the content of complex movies. Upon repeated viewings of extended movie clips, the activity patterns of individual neurons were highly repeatable, indicating that they were reliably driven by the movie content in the free viewing condition. Although the majority of neurons encountered in this region were nominally “face cells,” in that they responded more strongly to flashed images of faces compared with other object categories presented in a fixation task, in the context of movie viewing the responses of nearby neurons typically showed low correlations. Nonetheless, the neurons were systematically tuned to one feature, namely the size of faces and other body parts, showing that neurons in the AF face patch act not only as face cells but also as “size cells.” Since under ecologically relevant conditions variation in the size of the retinal image of faces is mainly driven by distance rather than by head size, our results demonstrate that AF neurons carry signals useful for extracting spatial relations among conspecifics in the animal’s vicinity. In the context of a natural viewing paradigm, the most conspicuous impact of this joint sensitivity to image category and scale is that AF neurons appear to act as “dis-

tance cells.” This observation is in harmony with the finding that neurons in the anterior STS that responded to locomotion were also sensitive to the proximity of human walking figures in addition to tuning for biological form and direction (Jellema et al., 2004). Moreover, neuropsychological observations in monkeys and human patients with lesions to the anterior STS revealed deficits in their ability to localize objects in the environment (Karnath, 2001), consistent with a role in spatial visual processing more commonly associated with the dorsal visual stream (Mishkin et al., 1983). Viewed in this light, it is reasonable to speculate that the anterior STS contains regions dedicated to spatial aspects of biological vision.

Role of AF within the face patch network

Face patch AF is one of many regions in the STS and ventral inferotemporal cortex that responds selectively to faces compared with other objects in the context of fMRI experiments (Tsao et al., 2003; Bell et al., 2009; Pinsk et al., 2009; Ku et al., 2011; Issa et al., 2013; Janssens et al., 2014). The question most

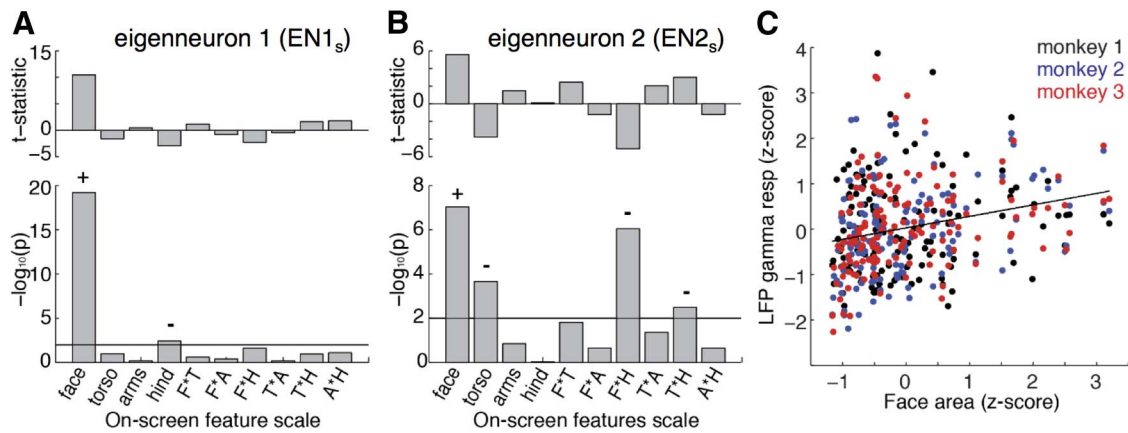


Figure 7. Impact of on-screen feature scale on neuronal activity. **A**, Outcome of multivariate regression analysis modeling the influence of four main effects of on-screen feature scale (face, torso, arms, and hindquarters) plus the six combinations of pairwise interaction effects (face X torso, face X arms, etc.) on activity of the first eigenneuron. The sign of statistically significant effects ($p < 0.01$) is indicated by + and – marks, as determined by the sign of the t statistic (t_{main}) in the case of main effects and by the sign of the product of $t_{\text{int}}(t_{\text{main1}} * t_{\text{main2}})$ for interaction effects. Positive effects are associated with enhanced eigenneuron activity, and negative effects with reduced activity. While faces showed a strong and positive influence on responses of EN1_s, other features showed either no influence or, in the case of hindquarters, a negative influence. **B**, Results of the same multivariate regression analysis applied to the second eigenneuron. Faces again showed a significant positive influence on EN2_s. Weaker effects, primarily negative, were observed for torso and two interaction terms. **C**, Positive relation between on-screen face scale and the gamma component of the LFP. F, face; T, torso; A, arms; H, hands.

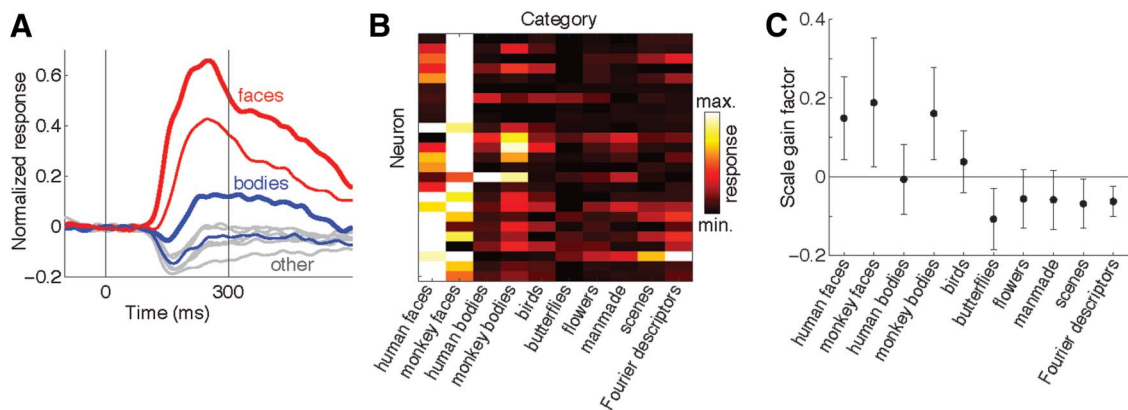


Figure 8. Category selectivity of AF neurons assessed with static images. **A**, Normalized population firing rate histograms for 25 AF neurons with positive responses. Red lines indicate mean population responses to monkey faces (thick line) and human faces (thin line). Blue lines indicate mean population responses to images of whole monkey bodies (thick line) and whole human bodies (thin line). Gray lines indicate responses to six other stimulus categories: birds, butterflies, flowers, man-made objects, scenes, and Fourier descriptors. **B**, Heat map showing mean responses of the same 25 neurons to the 10 stimulus categories. The color of each pixel indicates the mean response of single neuron evoked by all the stimuli within each category based on spikes counted within an 80–300 ms window after stimulus onset. **C**, Category dependency of image scale gain factor. The y-axis values indicate change in neuronal response in normalized (z-transformed) units per octave change in stimulus size. For categories with positive gain factors (such as faces and monkey bodies) larger stimuli induce stronger neuronal responses, whereas for categories with negative gain factors larger stimuli induce weaker responses.

directly raised by the current findings is whether other nodes of the face patch network display similar spatial sensitivities. While a definitive answer must await future recordings, the pattern of connections between face patches provides a means by which the properties seen in one node of the face network could be propagated to other nodes (Moeller et al., 2008). In the event that spatial sensitivity proves to be a general property of face-selective neurons throughout the temporal lobe, it will be of interest to determine through inactivation studies whether AF is the main source of spatial information, or alternatively whether it derives such input from other face patches. An alternative possibility, equally consistent with the currently available evidence, is that the spatial sensitivity we observed here might prove to be a distinct property of AF neurons. Such a result would point to a division of labor within the face network between regions specialized for the perception of face identity (culminating in the anterior medial

(AM) face patch; Leopold et al., 2006; Freiwald and Tsao, 2010) and separate regions dedicated to spatial aspects of face perception (including AF). The cluster of face patches in the posterior temporal lobe carry viewpoint-dependent selectivity for local features that could provide the building blocks for the view-invariant identity tuning seen in AM (Freiwald et al., 2009; Issa and DiCarlo, 2012). The same cluster of early face patches could convey low-level information about facial features to AF, where it could be integrated with the dorsal stream inputs that we consider next.

Integrating biological form with spatial context in AF

How might neurons in the AF face patch acquire their sensitivity to the spatial aspects of the visual scene that we observed here? Evidence from anatomical tracing studies indicates that it is the dorsal/fundus region of STS that receives diverse types of visual and nonvisual input, much of which is bidirectional. Projections

in the lateral convexity of the inferotemporal cortex (TEad; Saleem et al., 2000), and other regions of the STS fundus (Seltzer and Pandya, 1994), may provide ventral stream information about complex visual form, including biological actions. Such projections may provide input to the neurons long known to be sensitive to biological form and motion (Perrett et al., 1982; Desimone et al., 1984). Other projections from the posterior parietal cortex (Seltzer and Pandya, 1984, 1994; Cusick et al., 1995) may add dorsal stream spatial context to this form information. Projections from the overlying superior temporal gyrus may further imbue the AF face patch with multimodal properties (Seltzer and Pandya, 1994), as has been shown for neurons in the more broadly defined superior temporal polysensory area (Bruce et al., 1981). Finally, parahippocampal gyrus (Seltzer and Pandya, 1994; Suzuki and Amaral, 1994; Lavenex et al., 2002; Kondo et al., 2005) may convey information related to navigation and perceived spatial location (Bell et al., 2009; Kravitz et al., 2011; Kornblith et al., 2013). Together, these anatomical findings suggest that the anterior dorsal STS stands at a point of confluence between the dorsal and ventral visual streams.

Merits and challenges of naturalistic paradigms

The paradigm we used in this study represents a departure from the conventional investigation of face-selective neurons using flashed, static stimuli. The complex nature of the stimulus material presents both challenges and opportunities for discovering unexpected properties of brain function. The strategy taken here was to employ a highly data-driven approach to discover the principal modes of neural modulation across the population. This approach, combined with systematic comparison of each neuron's response patterns to the movies, led us to conclude that spatial scale was paramount in the response patterns of AF neurons. This finding was borne out by the striking correlation between facial scale and the response of the eigenneurons. At the same time, it is worth noting that many individual neurons displayed idiosyncratic response properties that appeared unrelated to apparent viewing distance, and despite their consistency across trials were not readily explained by any obvious feature of the movie content. It is possible that shedding light on the more enigmatic properties of these socially driven neurons might require still further strides in the direction of ethologically relevant physiological studies, such as telemetry recordings in freely moving animals.

References

- Barracough NE, Perrett DI (2011) From single cells to social perception. *Philos Trans R Soc Lond B Biol Sci* 366:1739–1752. [CrossRef Medline](#)
- Bell AH, Hadj-Bouziane F, Frihauf JB, Tootell RB, Ungerleider LG (2009) Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. *J Neurophysiol* 101:688–700. [CrossRef Medline](#)
- Bell AH, Malecek NJ, Morin EL, Hadj-Bouziane F, Tootell RB, Ungerleider LG (2011) Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity. *J Neurosci* 31:12229–12240. [CrossRef Medline](#)
- Bondar IV, Leopold DA, Richmond BJ, Victor JD, Logothetis NK (2009) Long-term stability of visual pattern selective responses of monkey temporal lobe neurons. *PLoS One* 4:e8222. [CrossRef Medline](#)
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436. [CrossRef Medline](#)
- Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369–384. [Medline](#)
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173. [CrossRef Medline](#)
- Cusick CG, Seltzer B, Cola M, Griggs E (1995) Chemoarchitectonics and corticocortical terminations within the superior temporal sulcus of the rhesus monkey: evidence for subdivisions of superior temporal polysensory cortex. *J Comp Neurol* 360:513–535. [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](#)
- 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](#)
- Furl N, Hadj-Bouziane F, Liu N, Averbeck BB, Ungerleider LG (2012) Dynamic and static facial expressions decoded from motion-sensitive areas in the macaque monkey. *J Neurosci* 32:15952–15962. [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](#)
- Ito M, Tamura H, Fujita I, Tanaka K (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73:218–226. [Medline](#)
- Janssens T, Zhu Q, Popivanov ID, Vanduffel W (2014) Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. *J Neurosci* 34:10156–10167. [CrossRef Medline](#)
- Jellema T, Maassen G, Perrett DI (2004) Single cell integration of animate form, motion and location in the superior temporal cortex of the macaque monkey. *Cereb Cortex* 14:781–790. [CrossRef Medline](#)
- Karnath HO (2001) New insights into the functions of the superior temporal cortex. *Nat Rev Neurosci* 2:568–576. [CrossRef Medline](#)
- Kondo H, Saleem KS, Price JL (2005) Differential connections of the perirhinal and parahippocampal cortex with the orbital and medial prefrontal networks in macaque monkeys. *J Comp Neurol* 493:479–509. [CrossRef Medline](#)
- Kornblith S, Cheng X, Ohayon S, Tsao DY (2013) A network for scene processing in the macaque temporal lobe. *Neuron* 79:766–781. [CrossRef Medline](#)
- Köteles K, De Mazière PA, Van Hulle M, Orban GA, Vogels R (2008) Coding of images of materials by macaque inferior temporal cortical neurons. *Eur J Neurosci* 27:466–482. [CrossRef Medline](#)
- Kravitz DJ, Saleem KS, Baker CI, Mishkin M (2011) A new neural framework for visuospatial processing. *Nat Rev Neurosci* 12:217–230. [CrossRef Medline](#)
- Ku SP, Tolias AS, Logothetis NK, Goense J (2011) fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron* 70:352–362. [CrossRef Medline](#)
- Lavenex P, Suzuki WA, Amaral DG (2002) Perirhinal and parahippocampal cortices of the macaque monkey: projections to the neocortex. *J Comp Neurol* 447:394–420. [CrossRef Medline](#)
- Leite FP, Tsao D, Vanduffel W, Fize D, Sasaki Y, Wald LL, Dale AM, Kwong KK, Orban GA, Rosen BR, Tootell RB, Mandeville JB (2002) Repeated fMRI using iron oxide contrast agent in awake, behaving macaques at 3 Tesla. *Neuroimage* 16:283–294. [CrossRef Medline](#)
- Leopold DA, Bondar IV, Giese MA (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442:572–575. [CrossRef Medline](#)
- McMahon DB, Olson CR (2009) Linearly additive shape and color signals in monkey inferotemporal cortex. *J Neurophysiol* 101:1867–1875. [CrossRef Medline](#)
- McMahon DB, Bondar IV, Afuwape OA, Ide DC, Leopold DA (2014a) One month in the life of a neuron: longitudinal single unit electrophysiology in the monkey visual system. *J Neurophysiol* 112:1748–1762. [CrossRef Medline](#)
- McMahon DB, Jones AP, Bondar IV, Leopold DA (2014b) Face-selective neurons maintain consistent visual responses across months. *Proc Natl Acad Sci U S A* 111:8251–8256. [CrossRef Medline](#)
- Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: two cortical pathways. *Trends Neurosci* 6:414–417. [CrossRef](#)
- Mitzdorf U (1987) Properties of the evoked potential generators: current

- source-density analysis of visually evoked potentials in the cat cortex. *Int J Neurosci* 33:33–59. [CrossRef Medline](#)
- 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](#)
- Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R (2005) Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309:951–954. [CrossRef Medline](#)
- Op De Beeck H, Vogels R (2000) Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426:505–518. [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](#)
- Pinsk MA, DeSimone K, Moore T, Gross CG, Kastner S (2005) Representations of faces and body parts in macaque temporal cortex: a functional MRI study. *Proc Natl Acad Sci U S A* 102:6996–7001. [CrossRef Medline](#)
- Pinsk MA, Arcaro M, Weiner KS, Kalkus JF, Inati SJ, Gross CG, Kastner S (2009) Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *J Neurophysiol* 101:2581–2600. [CrossRef Medline](#)
- Polosecki P, Moeller S, Schweers N, Romanski LM, Tsao DY, Freiwald WA (2013) Faces in motion: selectivity of macaque and human face processing areas for dynamic stimuli. *J Neurosci* 33:11768–11773. [CrossRef Medline](#)
- Russ BE, Leopold DA (2015) Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *Neuroimage* 109:84–94. [CrossRef Medline](#)
- Saleem KS, Suzuki W, Tanaka K, Hashikawa T (2000) Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. *J Neurosci* 20:5083–5101. [Medline](#)
- Seltzer B, Pandya DN (1984) Further observations on parieto-temporal connections in the rhesus monkey. *Exp Brain Res* 55:301–312. [Medline](#)
- Seltzer B, Pandya DN (1994) Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J Comp Neurol* 343:445–463. [CrossRef Medline](#)
- Shepherd SV, Steckenfinger SA, Hasson U, Ghazanfar AA (2010) Human-macaw gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr Biol* 20:649–656. [CrossRef Medline](#)
- Sripati AP, Olson CR (2010) Responses to compound objects in monkey inferotemporal cortex: the whole is equal to the sum of the discrete parts. *J Neurosci* 30:7948–7960. [CrossRef Medline](#)
- Suzuki WA, Amaral DG (1994) Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J Comp Neurol* 350:497–533. [CrossRef Medline](#)
- Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. *Nat Neurosci* 6:989–995. [CrossRef Medline](#)
- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670–674. [CrossRef Medline](#)
- Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci U S A* 105:19514–19519. [CrossRef Medline](#)