Behavioral/Systems/Cognitive

Precedence of the Eye Region in Neural Processing of Faces

Elias B. Issa and James J. DiCarlo

McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Functional magnetic resonance imaging (fMRI) has revealed multiple subregions in monkey inferior temporal cortex (IT) that are selective for images of faces over other objects. The earliest of these subregions, the posterior lateral face patch (PL), has not been studied previously at the neurophysiological level. Perhaps not surprisingly, we found that PL contains a high concentration of "face-selective" cells when tested with standard image sets comparable to those used previously to define the region at the level of fMRI. However, we here report that several different image sets and analytical approaches converge to show that nearly all face-selective PL cells are driven by the presence of a single eye in the context of a face outline. Most strikingly, images containing only an eye, even when incorrectly positioned in an outline, drove neurons nearly as well as full-face images, and face images lacking only this feature led to longer latency responses. Thus, bottom-up face processing is relatively local and linearly integrates features— consistent with parts-based models— grounding investigation of how the presence of a face is first inferred in the IT face processing hierarchy.

Introduction

Psychophysical work suggests that face processing is holistic in that it ultimately depends on all parts of the face (Tanaka and Farah, 1993). But computational work has shown that certain regions of the face are more reliable than others for distinguishing faces from nonface objects, and these informative regions often include one or both eyes (Viola and Jones, 2001; Ullman et al., 2002). Little is known presently about which strategy is instantiated neurally as features from simple parts to the whole face have been proposed to drive face-selective cells (Perrett et al., 1982; Desimone et al., 1984; Kobatake and Tanaka, 1994) or "face cells" that are found throughout primate inferior temporal cortex (IT) (Gross et al., 1972; Rolls, 1984; Tsao et al., 2006). Previously, the functional magnetic resonance imaging (fMRI)-based localization of patches of cortical tissue within IT that are highly enriched with face-selective cells (Tsao et al., 2003, 2008) has allowed more principled study of face processing, and neural recordings have revealed a buildup of face selectivity from the middle to the anterior IT face patches (Freiwald and Tsao, 2010). Such a hierarchical face processing strategy suggests that the visual features driving face-selective cells should be least complex yet most experimentally tractable in the posterior face patch and might be linearly recoverable in early, feedforward responses, which tend to be more linear than late responses (Brincat and Connor, 2006) yet contain sufficient information for decoding object category including faces (Hung et al., 2005). Here, we aimed to study the very first steps of face processing in IT—the earliest spikes elicited by neurons in the posterior face patch. We found that an overwhelming majority of cells were primarily driven by eye-like features placed within a region of the visual field [receptive field (RF)] just above the center of gaze, with tolerance to position and scale within that RF. In addition to eye-like features, image features in the face outline were a major contributor to initial responses, but image features in the region of the nose, mouth, and other eye made little contribution to the early response despite contributing to later activity. Our results demonstrate that face processing in the ventral stream begins with contralateral eye-like features in the context of a boundary (e.g., face outline) and does not depend on whole-face templates.

Materials and Methods

Animals and surgery

Two rhesus macaque monkeys (*Macaca mulatta*) weighing 6 kg (Monkey 1, female) and 7 kg (Monkey 2, male) were used in this study. Before behavioral training, a surgery using sterile technique was performed under general anesthesia to implant an MRI-compatible plastic head post in dental acrylic anchored to the skull using 16–20 ceramic screws (length, 5 mm; Thomas Recording). After fMRI scanning was completed, a second surgery was performed to place a plastic cylindrical recording chamber (19 mm inner diameter; Crist Instruments) over a craniotomy targeting the temporal lobe in the left hemisphere from the top of the skull (Monkey 1, Horsley–Clarke coordinates, +14 mm posterior-anterior, +21 mm medial-lateral, 8° medial-lateral angle; Monkey 2, +5 mm posterior-anterior, +23 mm medial-lateral, 8° medial-lateral angle). Angled (7.5 and 12°) or flat grids were used to reach the posterior, middle, and anterior face patches. All procedures were performed in compliance with National Institutes of Health guidelines and the standards of the MIT Committee on Animal Care and the American Physiological Society.

Received May 17, 2012; revised Sept. 10, 2012; accepted Sept. 16, 2012.

Author contributions: E.I. and J.D. designed research; E.I. performed research; E.I. analyzed data; E.I. and J.D. wrote the paper.

This work was supported by National Eye Institute Grant R01-EY014970, the McGovern Institute for Brain Research, and NIH National Research Service Award Postdoctoral Fellowship F32-EY019609. We thank H. Op de Beeck, A. Papanastassiou, P. Aparicio, J. Deutsch, and K. Schmidt for help with MRI and animal care, and B. Andken and C. Stawarz for help with experiment software.

The authors declare no competing financial interests.

Correspondence should be addressed to Elias Issa, McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 46-6157, Cambridge, MA 02139. E-mail: issa@mit.edu.

DOI:10.1523/JNEUROSCI.2391-12.2012

Copyright © 2012 the authors 0270-6474/12/3216666-17$15.00/0

Behavioral training

Following head-post implant surgery, subjects were trained to sit in a sphinx position (Vanduffel et al., 2001) in a custom chair compatible with our on-site horizontal bore scanner. Custom software (MWorks, Mac based) was used for stimulus display and data acquisition in all behavioral training, fMRI, and physiology experiments. Eye position was monitored by tracking the position of the pupil using a camera-based system (MRI, Iscan; physiology, Eyelink II). At the start of each training session, subjects performed an eye-tracking calibration task by saccading to a range of spatial targets and maintaining fixation for 400 ms. The best-fitting least squares linear model (four total parameters for horizontal/vertical offset and gain) was used for the remainder of the session in most cases; however, if drift was noticed over the course of the session, animals were recalibrated (DiCarlo and Maunsell, 2005). Training before fMRI scan sessions took place in a mock scanner. Subjects were trained to fixate in a 4° window for as long as possible by decreasing the interreward interval linearly with time (Vanduffel et al., 2001) (interreward interval starts at 2.4 s and decreases by 200 ms on every reward) such that reward per unit time increased exponentially up to the imposed limit (minimum interreward interval, 1.2 s), and a large reward was given at the end of scanner runs for good overall performance. Blinks defined as deviations outside the fixation window shorter than 200 ms in duration were allowed during extended fixations. After 2–3 months of training, subjects were able to sustain fixation >90% of the time during the 5 min period of each scanner run. For physiology experiments, subjects were required to fixate for shorter intervals than in the scanner (3 s physiology trials vs 5 min scanner runs). Given their prior training in the scanner, subjects quickly adjusted to fixating during individual physiology trials, but no blinks were allowed during these shorter fixations.

Images and presentation times

All images for experiments were generated at 400 × 400 pixel resolution with 256-level grayscale. Gray levels for each image were subtracted by the mean and normalized by the SD across the image. In physiology experiments, images were drawn from four categories (monkey faces, headless monkey bodies, objects, and indoor scenes; 10 exemplars each). Images were presented on a 24 inch cathode ray tube monitor (1900 × 1200 pixels, 85 Hz refresh rate, 19.0 × 12.1 inch viewable area, 48 × 31° of visual angle subtended; Sony GDM-FW900) positioned 55 cm in front of the animal. Each image was placed on a different instance of a pink noise background matched for the mean power spectrum across all of the images. This precaution was taken instead of using gray backgrounds to ensure equal luminance and contrast across the image for objects of varying aspect ratio (i.e., face and scene images filled a larger portion of the bounding box compared to bodies, which were elongated in one direction). Rapid serial visual presentation (RSVP) was used with 100 ms on and 100 ms off durations, and 15 images were presented per trial with a 400 ms pretrial fixation period (total fixation duration, 3.4 s); the first image in a trial was not considered. Images tested in RSVP mode were presented at a size of 6° and included the general screen set (faces, bodies, objects, and places; 10 exemplars each), face part combinations (left eye, right eye, nose, mouth, and outline were present, absent, or varied in position), and occluded faces (horizontal or vertical bar at five positions). Although faces from 10 individuals were shown in the screen set (see Fig. 1), all follow-up image sets (reverse correlation, face parts combinations, occluded faces, and retinal mapping) used manipulations of the same monkey face image for testing all sites (see Figs. 2–12). Five repetitions were collected for the screen set (Fig. 1), 7 repetitions per spatial position for reverse correlation mapping (see Figs. 2–4, 6, 7), 2 repetitions per spatial position for receptive field mapping (see Fig. 7), 10 repetitions for face parts combinations (see Figs. 8–12), and 15 repetitions for occluded faces (see Figs. 10–12).

Following initial testing with the general screen set, most sites were further characterized using a reverse correlation mapping procedure. The position of a Gaussian window (FWHM, 1.5°; $\sigma = 0.64°$) was randomly varied every four frames (47 ms) across a base monkey face image (same exemplar used for all sites) such that at any given moment only a small fragment of the face was visible (Gosselin and Schyns, 2001) (see Fig. 2*A*). The position of the window was uniformly sampled across the 6° extent of the image, and at least 1000 samples were collected (28 samples/deg^2) at 20 samples per second over the course of 1 min of testing at each site. On a given trial, up to 62 positions were tested (corresponding to 3 s fixation), and the first window position tested (first four frames) in any trial was not considered. Because of the rapid nature of presentation, the fragments could merge perceptually and produce the impression that a face was present even though information was only presented locally. Choosing the size of the window reflected a trade-off between eliciting strong responses (larger window) or gaining resolution (small window). We chose a window that covered approximately a quarter of the image (FWHM, 1.5°; image size, 6°) because this was a comparable scale to the parts (e.g., eye, nose, mouth) typically identified in faces and was a reasonable size for features that might activate sites in posterior IT. Using smaller windows or larger windows in control tests did not change the main finding.

In a subset of sites, reverse correlation mapping was repeated for altered versions of the 6° base image: 3° size, 12° size, contrast limited adaptive histogram equalized (CLAHE), cartoon, contrast reversed, horizontally shifted 3°, and inverted. Cartoon and CLAHE faces were meant to equate the local contrast and spatial frequency of the features in the face, and these were only a first-pass attempt at determining the robustness of face-component preferences to different low-level factors. CLAHE faces were generated by equating luminance histograms in local regions (50 × 50 pixels = 0.75 × 0.75°) of the image and combining neighboring regions using bilinear interpolation to eliminate artificial boundaries (Matlab function adapthisteq). This resulted in a uniform distribution of luminance values in each 50 × 50 region of pixels, which implicitly equalizes local contrast and spatial frequency across the image. For cartoon faces, the image was first decomposed into a cartoon (bounded variation) plus texture (oscillatory) component (Buades et al., 2010) (texture scale, 4 pixels; C code publicly available at http://www.ipol.im./pub/algo/blmv_nonlinear_cartoon_texture_decomposition/) followed by mapping of clustered regions of the luminance histogram (*k*-means clustering) into a compressed gray scale from 256 to 3 gray levels (posterization). The goal of this transform was to delineate part boundaries in the image and remove texture edges. Setting the luminance constant within each segment or "part" effectively sets contrast and spatial frequency of the interior to zero. The only remaining contrast is on the part boundaries, but this contrast is equalized across all edges by the posterization. Although we used a nonlinear cartoon plus texture decomposition, similar results could be obtained using a linear low- plus high-frequency filter pair or by using median filtering, which blurs edges less than parametric filters.

Retinotopic receptive field mapping was performed in a subset of sites by flashing a 3° face (50 ms on, 150 ms off) centered at a random position chosen on a polar grid spanning −6 to 6° of visual angle (300 samples collected). This sampling strategy led to denser samples near the fovea, where resolution is highest, and sparser samples in the periphery.

To manipulate the presence or absence of face parts, we first hand segmented the parts (eye, nose, and mouth). A synthetic face was created by starting with the face outline filled with pink noise and then blending in each part using a Gaussian window for the α (transparency) value. This approach was taken—rather than reducing further to cartoons or line drawings, which allow more parametric control—since we had no a priori guess as to what features in face images drive face-selective sites. For example, conceptualizing the "eye" as a circular outline in a cartoon drawing ignores the statistics of features such as the light/dark texture boundary surrounding the eye including the brow ridge, the junction formed by the nose bridge and the eye, and the interior details of the iris, pupil, and cornea, which could all be important in driving responses. Using fragments from a natural face image helps preserve these features.

Face images (size, 6°) were occluded using horizontal or vertical gray bars (horizontal bar, 6 × 1.2; vertical bar, 1.2 × 6°) centered at one of five equally spaced positions (−2.4, −1.2, 0, +1.2, or +2.4°) in the horizontal or vertical direction.

X ray-based electrode localization

All recordings were conducted under the guidance of a microfocal stereo x-ray system developed previously in our lab (Cox et al., 2008). Monkeys

were fitted with a plastic frame (3 × 4 cm) containing six brass fiducial markers (1 mm diameter). The frame was positioned over the lateral aspect of the temporal lobe using a plastic arm anchored in the dental acrylic implant. The fiducial markers formed a fixed 3D skull-based coordinate system for registering all physiological recordings. At each recording site, two x rays were taken simultaneously at near orthogonal angles. In each x-ray image, the 2D coordinates of the fiducials (automatically extracted using a fast radial symmetry transform) (Loy and Zelinsky, 2003) and the position of the electrode tip were determined. Given the known system geometry (spacing and angle between x-ray sources and detectors recovered from imaging a calibration object), stereophotogrammetric techniques were used to reconstruct the 3D position of the electrode tip in the common fiducial reference frame allowing coregistration of sites within and across days. To coregister recording positions to MRI volumes, an anatomical scan was obtained with the reference frame attached. The frame was visualized in MRI through four copper sulfate (Cu_2SO_4)-filled wells (2 mm diameter) whose positions relative to the brass fiducials were determined previously (off-site micro-computed tomography of plastic frame at 25 u resolution; Micro Photonics). A best-fitting affine transform between x-ray and MRI coordinate systems was then used to project physiology sites onto structural and functional MRI volumes.

Physiological recordings

Multiunit activity (MUA) was recorded using glass-coated tungsten microelectrodes (impedance, 0.5 MΩ; outer diameter, 310 u; Alpha Omega). A water-based hydraulic microdrive (Kopf Instruments) was used to lower electrodes through a 26 gauge stainless-steel guide tube inserted into the brain (10–20 mm) and held by a plastic grid inside the recording chamber. The signal was passed through a high-impedance headstage (gain, 10), amplified (gain, 10^4; Bak Electronics), and digitally filtered (300 Hz to 4 kHz passband; Krohn-Hite) before applying a threshold to obtain MUA event counts. MUA event detection thresholds were determined at the beginning of each session by the experimenter and were set from one to two SDs above baseline. The same threshold was used throughout the remainder of the session. The raw electrode signal (1 Hz to 4 kHz, 8 kHz sampling rate), horizontal and vertical eye position trace (1 kHz sampling rate), and vertical refresh signal (1 kHz sampling rate) were stored for later use off-line. To ensure accurate stimulus locking, spikes were aligned to the onset of the first frame/vertical refresh after a display command was issued. Since this work was part of a larger mapping study, recordings were made systematically at 300 u intervals throughout IT such that sampling was unbiased for the screen set of images that was tested on all sites. Sites that were visually driven were further tested with other image sets.

Individual single units were sorted off-line using an automated clustering algorithm (Quiroga et al., 2004). The filtered signal (300 to 3500 Hz) was first thresholded to obtain putative spike events. Spike waveforms were then decomposed into wavelet features before superparamagnetic clustering. In some cases, the spike detection threshold or the temperature used by the clustering algorithm were adjusted manually during *post hoc* inspection. Waveforms whose peak-to-peak amplitude was at least five times that of the SD of the background noise were considered as single units [signal-to-noise ratio (SNR) $\geq 20 * \log_{10}(5) = 13.97$ dB], and a total of 257 single units passing this criterion were recorded in identified area posterior lateral face patch (PL) using the general screen set (see Fig. 1*C*). Smaller subsets of single units were analyzed in follow-up image tests as long as they showed some visual drive to an image in those sets (≥1× SEM above baseline) (see Figs. 6*A*, 11) or, in the case of reverse correlation mapping, as long as predictions of linear weight maps were significant (p value of split-halves cross-validation of <0.05; see Subspace reverse correlation, below). All single-unit analyses were confined to be within a single image set since no effort was made online to ensure single-unit isolation across image sets.

Analysis

Physiological definition of posterior face patch. Responses on the general screen set were pooled into target (face, 10 exemplars, five repetitions, 50 total trials) or distractor (body/object/place, three categories, 10 exemplars, five repetitions each, 150 total trials) classes. These two response distributions were then compared using a d' measure for face selectivity computed from spike counts in a 50–200 ms poststimulus window for each site:

$$d' = \frac{(\mu_F - \mu_{BOP})}{\sqrt{\frac{\sigma_F^2 + \sigma_{BOP}^2}{2}}}.$$

The spatial profile of selectivity in native 3D brain tissue space was fit with a graded sphere model where selectivity begins at a baseline level β outside the sphere and increases linearly with slope a toward the center of the sphere. The diameter of the spherical region was varied from 1.5 to 10 mm, and candidate centers were tested within a 5 mm radius of the predetermined fMRI-based center of the patch. For each position and radius tested, the slope α and the baseline β were fit to sites within a 10 mm region (minimum of 15 sites required). The graded sphere producing the highest correlation (Pearson) with the actual selectivity profile was used as the physiological definition of the center and boundary of the posterior face patch. The diameter of the best-fit sphere was 3.75 mm in Monkey 1 and 5.0 mm in Monkey 2. For Figure 1, *B* and *D*, a split-halves approach was used to avoid selection bias. A spherical region was fit using half of the physiology sites, and face selectivity metrics were computed on the remaining half of the sites, which were held out of the localization procedure. For all other analyses on independent image sets, the best-fitting sphere for all of the sites was used.

Visual response screen. Multiunit sites were considered visually driven if their mean response in a 60–160 ms window was >2 × SEM above baseline for at least one of the four categories tested (faces, bodies, objects, or places). All sites passing this general screen were included for reverse correlation analysis and other follow-up image tests as no other response screens were applied.

Firing rate. All firing rates used in the main analyses were computed in a 40 ms window beginning at the response latency of each site. A fixed 0–50 ms window (locked to image onset) was used to compute baseline firing rates for a particular image or category and was subtracted from raw firing rates to obtain driven firing rate measures.

Latency. Poststimulus time histograms (PSTHs) were binned at 1 ms resolution and baseline subtracted. PSTHs were not smoothed before analyses. Response latency was computed using a change point estimation algorithm (Friedman and Priebe, 1998). The change point was defined as the point that minimized the error of a piecewise linear fit to the knee of the cumulative PSTH. The search for the optimal latency was bounded below at 50 ms and bounded above by the time of the peak response. The SE of latency estimates was computed using bootstrap resampling ($n = 15$ iterations).

Face selectivity index. The face selectivity index (FSI) was computed as follows:

$$\mathrm{FSI} = \frac{F - BOP}{|F| + |BOP|},$$

where F is the mean response to all face images ($n = 50$; 10 exemplars times five repetitions), and BOP is the mean response to all other images ($n = 150$; 30 exemplars times five repetitions). This measure was used for comparison to previous work (Tsao et al., 2006; Freiwald and Tsao, 2010). Similar results were obtained using a d' measure (see above), which does not pin at −1 or 1 and has a more continuous range of magnitudes (Ohayon et al., 2012).

Subspace reverse correlation. The neural response was modeled as the sum of the weighted contribution of each part of the image over time and was estimated using standard reverse correlation techniques (see Figs. 2–4, 6, 7). Linear weights were in units of spikes per second and were determined at 0.5° spatial resolution spanning −3 to 3° (12 × 12 = 144 spatial bins) and at 20 ms temporal resolution from 0 to 200 ms (20 temporal bins). On average, seven samples were obtained per spatial bin (1000 random positions tested/144 bins). This linear system of equations (1000 equations in 144 unknowns in each 20 ms time bin) can be solved in the least squares sense (find weights w for the stimulus matrix X and

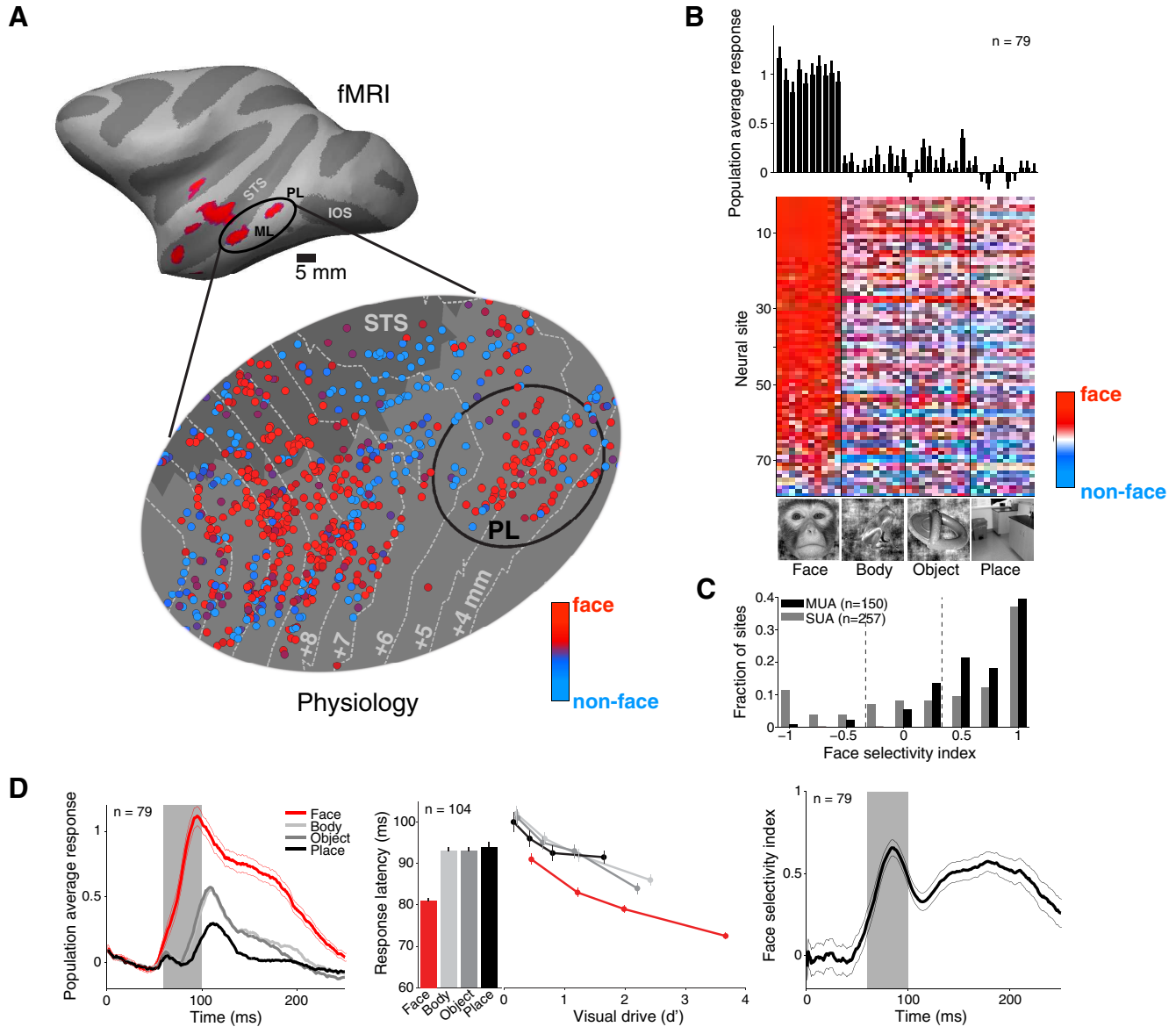

Figure 1. Face selectivity in PL. ***A***, Top, PL is the most posterior patch in the inferior temporal lobe demonstrating face selectivity under fMRI; ML is also labeled for reference. Data are shown for Monkey 1. fMRI in Monkey 2 also revealed the posterior face patch bilaterally. Bottom, Physiological map of multiunit recordings in IT reveals a posterior cluster of face-selective sites corresponding to PL as identified in fMRI (sites colored by neuronal response d' for faces vs the mean across bodies, objects, and places; site positions and physiologically defined boundary of PL, at its widest extent in 3D, were projected in native 3D space to the 2D boundary between gray and white matter before flattening of the 2D cortical patch; dashed contour lines are projected Horsely–Clarke AP coordinates). STS, Superior temporal sulcus; IOS, inferior occipital sulcus. ***B***, Responses of sites in PL to the standard screen set of faces, bodies, objects, and places (10 exemplars per category). Top, Individual site responses were averaged to generate the population response, which is shown rescaled by the average response to faces. Bottom, Sites were first rescaled by their maximum image response and rank-ordered according to their mean response across face exemplars. ***C***, Comparison of multiunit and single-unit face selectivity distributions (FSI; see Materials and Methods). Antipreferring face sites (FSI < 0) were more common in single units than in multiunits. ***D***, Left, Face images drove a strong early response compared to bodies, objects, and places (gray box denotes early response window of 60–100 ms; population-average response shown rescaled by the median response to the whole face at 100 ms). Middle, PL responses to faces had significantly shorter latencies than responses to bodies, objects, and places at all levels of visual drive (measured as d' of response above baseline; this analysis required between-site comparisons of latencies for responses with matched visual drive). Given that higher firing rates lead to shorter latencies slope, approximately -5 ms per unit of d'), this analysis demonstrates shorter latencies for faces compared to other object categories independent of overall visual drive. Right, The FSI peaked within the first 40 ms of the response.

the response y in $Xw = y$ such that $\|Xw - y\|^2$ is minimized) using the normal equation $w = (X^TX)^{-1}X^Ty$, where the stimulus autocorrelation matrix X^TX removes (whitens) spatial correlations induced by the Gaussian window across bins (DiCarlo and Johnson, 1999). This normalization, however, divides by near-zero values for off-diagonal elements, which can amplify noise. To limit the effects of noise, we performed ridge regression using Tikhonov regularization by introducing a diagonal regularization matrix (cost function $= \|Xw - y\|^2 + \alpha^2 * \|Iw\|^2$) that encourages smaller weights: $w_{\text{reg}} = (X^TX + \Gamma^T\Gamma)^{-1}X^Ty$, where $\Gamma = \alpha I$.

Enforcing this prior leads to solutions w_{reg} with smaller L_2 norms. The optimal regularization factor a was varied to minimize fit error (on a holdout set) using a cross-validation approach (split halves averaged across 10 splits; α varied from 0.001 to 10,000).

Goodness of fit of the derived linear model was computed by comparing the maximum possible explainable variance of neural responses (reproducible variance not due to noise; determined by correlation between repeated presentations in a spatial bin) to the variance explained by model fits under split-halves cross-validation (weights fit from half of the

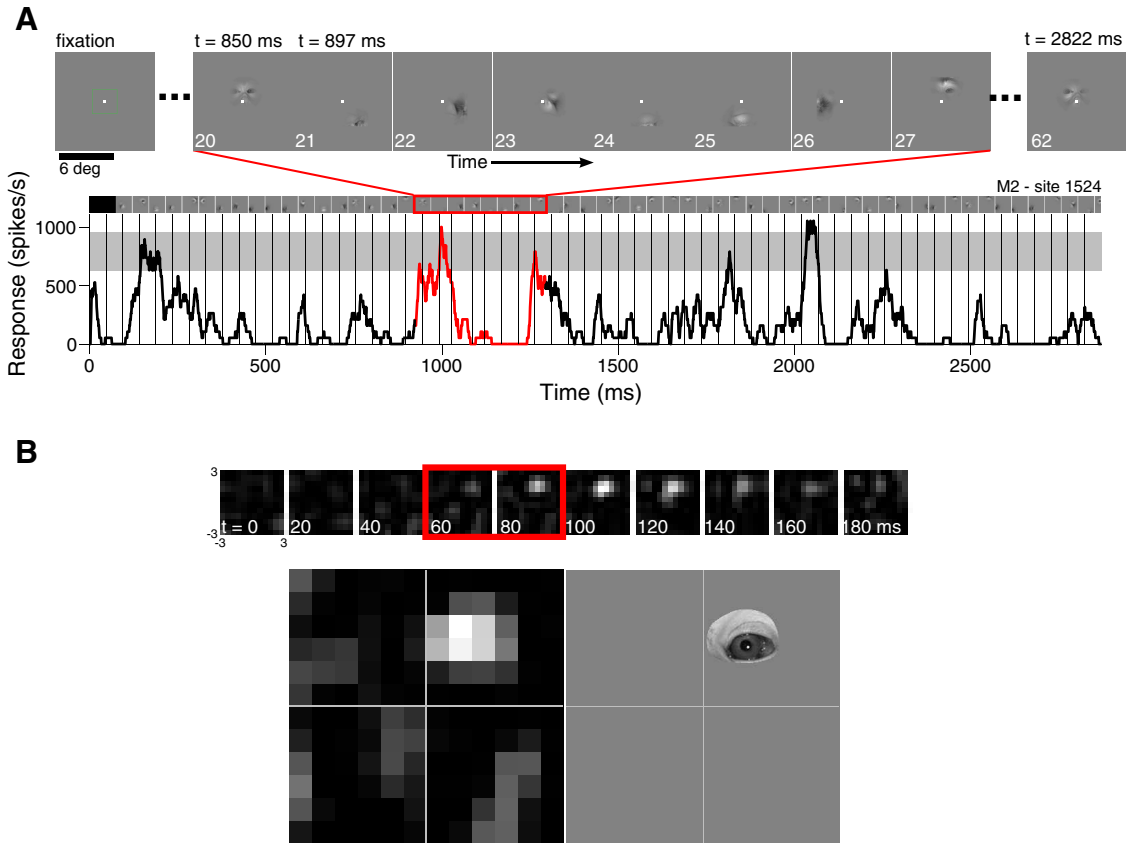

Figure 2. Face reverse correlation mapping. ***A***, Gaussian windowed face fragments were presented in rapid succession (every 47 ms or 21 Hz) while the monkey fixated a central fixation square (white) for up to 3 s. The example site (bottom row) responded strongly during frames when the interior of the eye was present. (Notice the strong response to the first and last frame in the red highlighted region. Frames are offset by the site's overall latency, 76 ms, to aid visual alignment with responses. The gray region denotes the response level in the screen set to the whole face). ***B***, Top, Reverse correlation analysis was performed to determine the linear response weighting of 0.5 × 0.5° regions of the face in 20 ms response bins. *Z*-scored weights are plotted. Bottom left, The spatial kernels in the early response window (60–100 ms; red rectangle) were averaged to generate the weighting map for the early response. Bottom right, The image fragment corresponding to the weighting region at half-height was extracted.

data were used to predict responses in the withheld half of the data, and correlations were averaged across five splits). In general, goodness-of-fit values were quite high (median $r_{\text{explainable}} = 0.52$, $r_{\text{predict split}} = 0.46$, $r_{\text{goodness-of-fit}} = 0.87$, $n = 111$ sites).

To estimate the mean and SD of weights expected under a null model with the same distribution of firing rates, we randomly paired stimuli and responses and recomputed linear spatiotemporal response fields. This shuffle control generated a noise distribution for weight values in each time bin. A normalized d' measure of weight strength was computed by subtracting the mean and dividing by the SD of the noise distribution in each time bin. On average, reverse correlation maps yielded 15 (of 144 possible) bins with values >1.5 SDs above baseline, which corresponds to a 3.75 deg^2 region ($n = 111$ sites; one site yielded no significant bins). Individual site d' values were averaged to generate population response fields (see Figs. 3, 6, 7) (similar results were obtained by averaging the raw weights).

The horizontal and vertical centers of reverse correlation maps were computed as the centers of mass of weights at least 1.5 SDs above shuffle controls. Centers were computed in the time bin with the highest explainable variance (i.e., most reliable time bin). Size was computed as the number of bins with a response >50% of the maximum (area at half-height) and scaled by 0.25 deg^2/bin. The uncertainty in position and size estimates was estimated using Monte Carlo simulation ($n = 30$ repeats) where Gaussian noise (SD estimated from shuffle controls) was added to weight maps before recomputing center position and size. Errors in the estimates of the center of weight maps for each site averaged 0.25°, and errors in the estimates of the size of weight maps averaged 1.2° (position, $x = 1.0, y = 1.3 \pm 0.25$° from center of 6° face image; size, 3.3 ± 1.2 deg^2; $n = 111$ sites).

Linear fit of response to wholes using responses to the parts. The response to each image was treated as a linear weighting of the left eye, right eye, nose, mouth, and outline yielding five parameters that were used to fit the response to 16 stimuli total ($2^4 = 16$ images containing all possible subsets of right eye, left eye, nose, and mouth with the outline held fixed) (see Fig. 8*A*). Parameters were determined using the normal equations to obtain a least squares solution to the linear regression problem:

$$y = \sum_{i=1}^{5} w_i I_i,$$

where the indicator function I_i is zero if the part is absent, and one if the part is present. Before computing linear fits, firing rates for each site were normalized by the average magnitude of responses. This was done to ensure that derived weight values were comparable in magnitude across sites. The weight of the outline corresponds to the offset term since the outline was present in all 16 tested images. This estimated weight was similar to the response of the outline when presented alone and did not reflect an effect of baseline since baseline was always subtracted before analyses.

Retinal receptive fields. Receptive field maps were constructed by binning responses at 1° resolution in a rectangular grid spanning −6 to 6° in 20 ms time bins (12 × 12 × 10 bins) (see Fig. 7*A*). The time bin with the maximum response was used to determine the center and size of receptive fields. All spatial positions evoking a response >1.5 SDs above baseline were used to compute the centroid of retinal receptive fields. Size was computed as the number of bins with a response >50% of the maximum (area at half-height) scaled by 1 deg^2/bin.

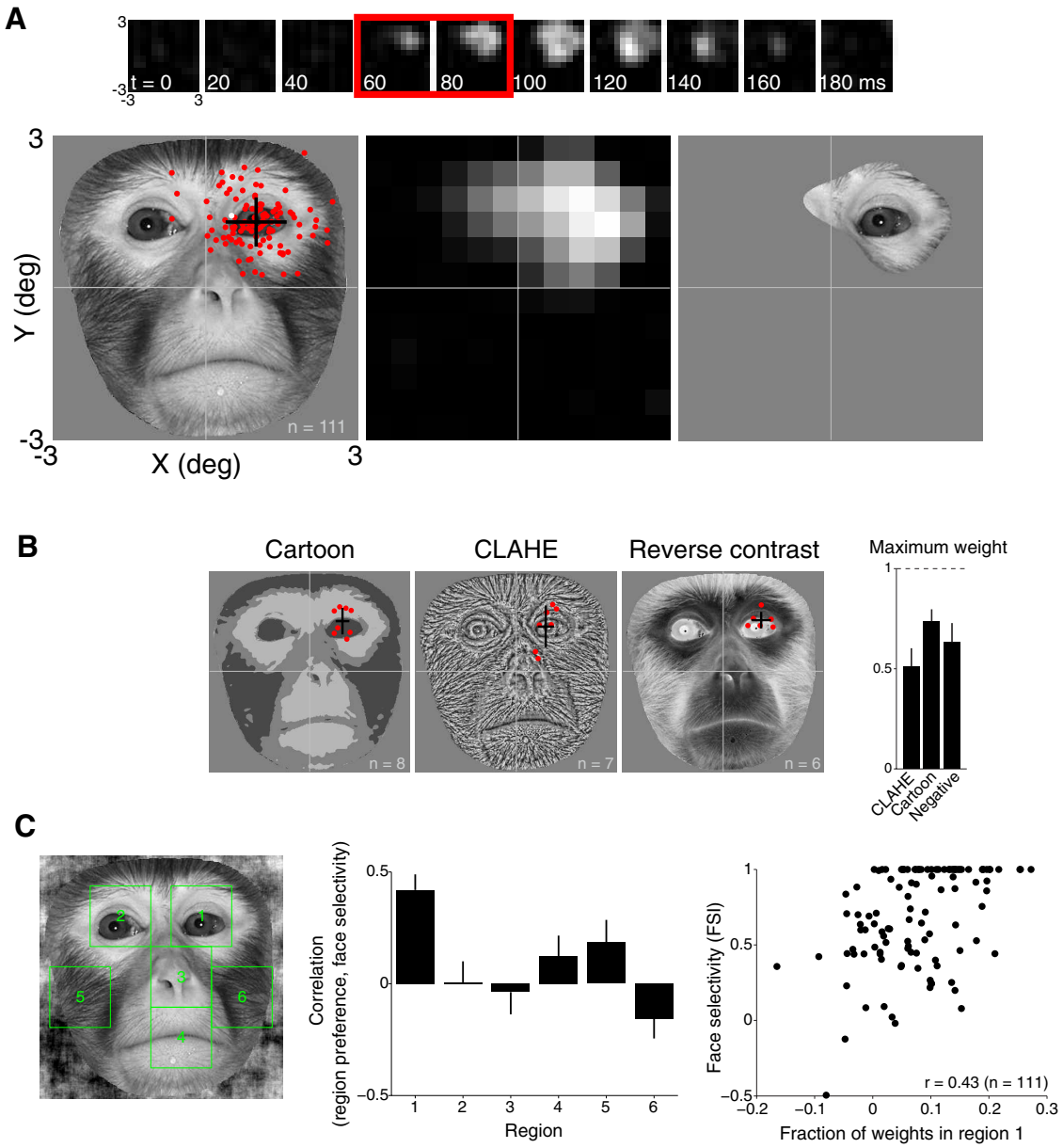

Figure 3. Preference for the eye region in PL. ***A***, In a majority of sites, weights in the early response window (60–100 ms; red outline) were centered in the top contralateral quadrant (left; dots reflect centroid of preferred region for individual sites; crosshair reflects mean and SD of *x*, *y* centers across sites; white dot is the center of the preferred region from the example site in Fig. 2). The corresponding face fragment (50% contour of the population response that was obtained by averaging *z*-scored weights across sites) was centered on the eye (middle, right). ***B***, The eye region was the most strongly weighted part of the face in cartoon (left), CLAHE (middle), and reversed-contrast faces (right). Peak response weights for these manipulated images are plotted relative to the median of the maximum weights for the original face image (far right). ***C***, The face selectivity index was correlated with the fraction of weights in the eye region in reverse correlation maps (right) but was not correlated with the fraction of weights in other regions of the face (middle). Six nonoverlapping regions across the face were tested (left).

Noise-adjusted correlation. Correlations between individual site responses and the sum of parts responses in Figure 8*C* were adjusted for noise in firing rates using the following:

$$r_{\text{normalized}} = \frac{r_{\text{explained}}}{r_{\text{explainable}}} = \frac{r_{\text{site, model}}}{\sqrt{r_{\text{site, site}} \cdot r_{\text{model, model}}}},$$

where $r_{\text{site, model}}$ is the raw correlation (e.g., correlation between the response to the whole and the sum of the responses to the parts) and $r_{\text{site, site}}$ or $r_{\text{model, model}}$ is the correlation between mean firing rates or sum of the parts responses on half of the trials and the withheld half of trials (averaged across 30 draws). The amount of agreement between different trial splits of the data provided an upper bound on explainable variance. Since in both the single site, $r_{\text{site, site}}$, and model estimates, $r_{\text{model, model}}$, only half of the data were used, the Spearman–Brown correction was applied according to $\hat{r}_n = 2 * r_{n/2}/(1 + |r_{n/2}|)$ for comparison to $r_{\text{site,model}}$, which used the whole dataset. Our rationale for using a noise-adjusted correlation measure for goodness of fit is to account for variance in the neural response that arises from noise and hence cannot be predicted by any model. The remaining variance driven by external stimuli is the benchmark for a model, and this ceiling is reflected by reproducibility across trials (DiCarlo and Johnson, 1999, their Appendix B; Op de Beeck et al., 2008, their supplemental material).

Statistics

Correlations were computed using a Spearman's rank correlation coefficient unless noted otherwise. A Wilcoxon rank-sum test was used for all statistical testing of size differences and was considered significant at the $p < 0.01$ level. With a few noted exceptions, all error bars and error margins reflect ±1 SEM.

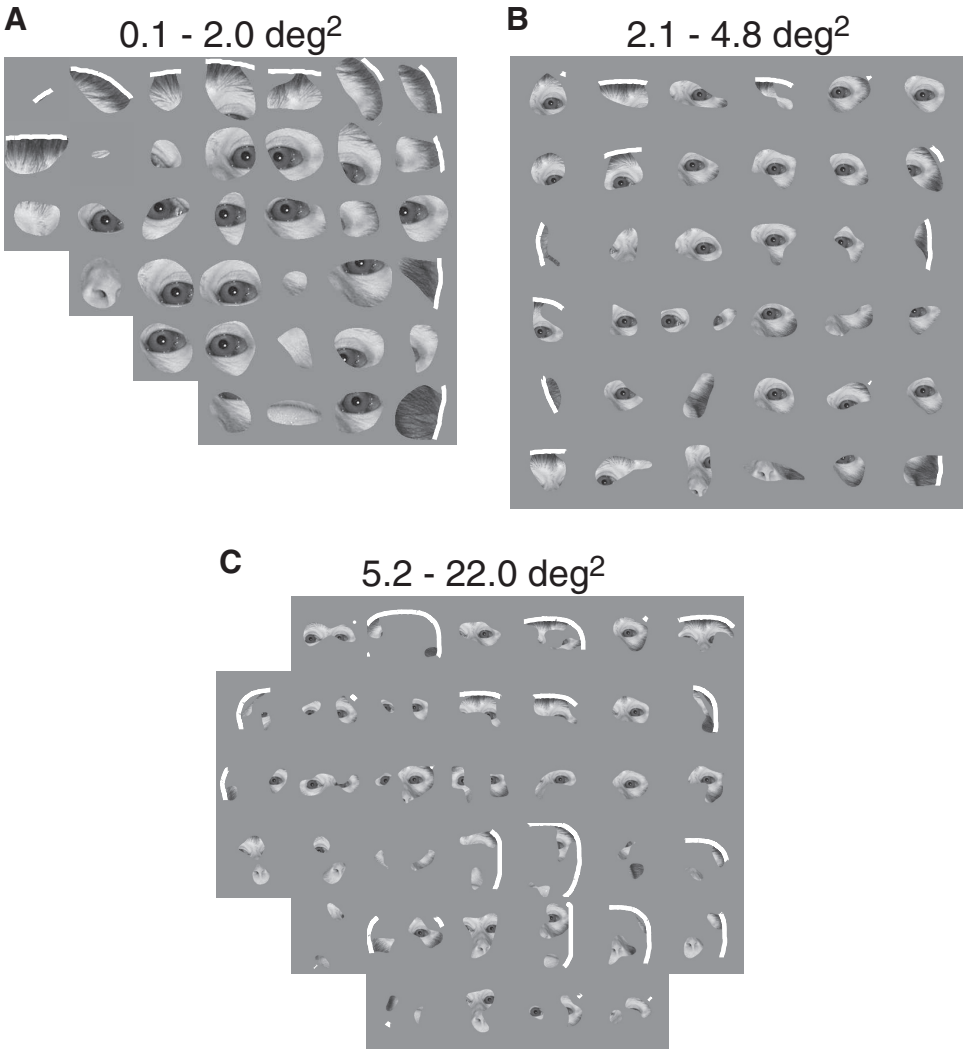

Figure 4. Preferred face fragments of individual sites in PL. Face fragments (50% regions) for individual sites are grouped based on their size and ordered by their relative position for ease of display. ***A***, ***B***, For sites whose preferred regions were small or intermediate in size, the right eye was the most common feature. ***C***, Larger regions tended to encompass both eyes. For sites with multiple half-height regions, only the two largest are plotted, and most sites had only one half-height region (***A***, ***B***). White outlines mark the edge of the face outline for reference, and features without a white outline lie toward the interior of the face.

Images and presentation times: fMRI

All images were generated at 400 × 400 pixel resolution with 256-level grayscale. Gray levels for each image were subtracted by the mean and normalized by the SD across the image. For fMRI scanning in Monkey 2 (for a description of fMRI scanning in Monkey 1, see Op de Beeck et al., 2008), images were drawn from five categories (monkey faces, human faces, headless monkey bodies, objects, and scenes; 20 exemplars each). Images were object cutouts placed on a white noise background, and we created scramble controls of all images using a texture synthesis algorithm described previously (Portilla and Simoncelli, 2000; Rust and DiCarlo, 2010) (Matlab code publicly available at http://www.cns.nyu.edu/~lcv/texture/). Images were rear projected (1024 × 768 pixels; 75 Hz refresh rate; Hitachi CP-X1200) onto a screen positioned 117 cm in front of the animal (15.0 × 11.25 inch viewable area, 19 × 14° of visual angle subtended) at 12° size for 800 ms on and 200 ms off in a block design (block length, 32.5 s; 33 images shown). Each run included 12 blocks where the first and last blocks were fixation blocks (gray screen), and the middle 10 blocks alternated between intact and scrambled images drawn randomly without replacement from the five categories such that exactly one repetition was collected for each intact/scrambled image category in a single run.

fMRI scanning

Functional scanning was conducted at the Martinos Center for Biomedical Imaging at Massachusetts General Hospital (Monkey 1) in a previous study (Op de Beeck et al., 2008) and at the Martinos Center in the McGovern Institute for Brain Research (Monkey 2) in the present study in a horizontal bore 3 Tesla Siemens Tim Trio magnet using a 4 inch saddle-shaped surface coil (receive only) positioned on top of the head. Before scanning, a monocrystalline iron oxide nanoparticle contrast agent (12.7 mg/ml at 8–10 mg/kg; Center for Molecular Imaging Research, Massachusetts General Hospital, Boston, MA) was injected intravenously in the saphenous vein (Vanduffel et al., 2001). To prevent iron buildup, an iron chelator (Desferal; Novartis International) was injected subcutaneously on a monthly basis (Leite et al., 2002), and blood levels of iron remained normal during quarterly roundup exams. At the beginning of each session before functional scans, the magnet was shimmed to adjust for magnetic field inhomogeneities, a localizer scan was used to position slices for complete coverage of the ventral visual processing stream (occipital and temporal lobes; only partial coverage of frontal lobe), the magnet was shimmed a second time, and finally a field map (two gradient echos at different echo times) was collected to measure geometrical distortions in the B_0 field for off-line dewarping of func-

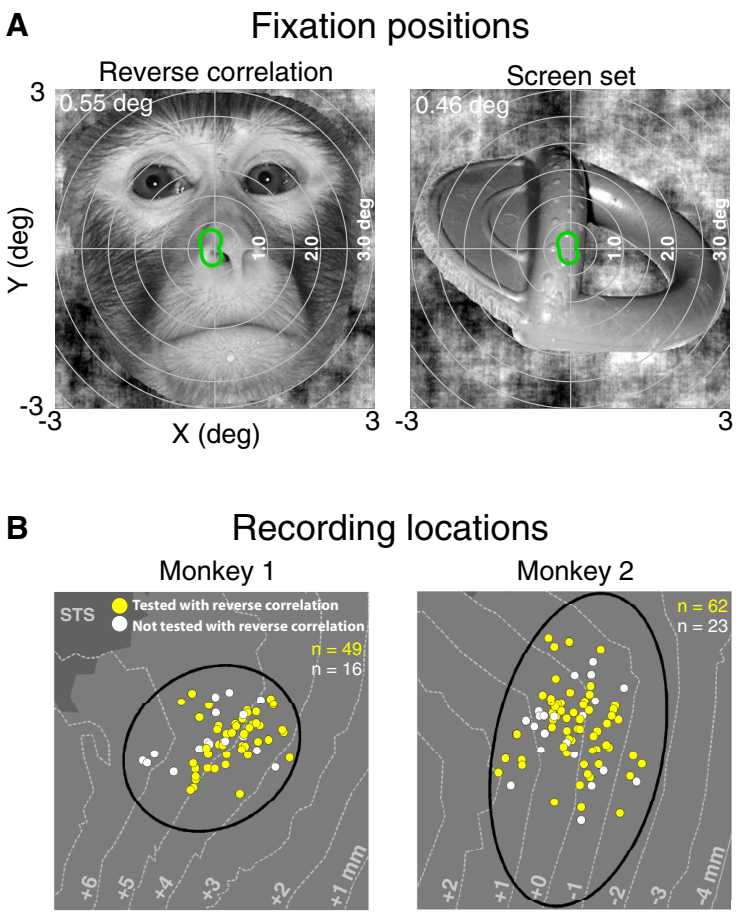

Figure 5. Eye fixations and recording locations for reverse correlation mapping. ***A***, Fixation positions during testing with the general screen set and during reverse correlation mapping. Contours represent 90th percentile of fixation positions along each radial direction (sampled at 15° intervals). Polar grids are plotted at 0.5° spacing. ***B***, Recording maps in each monkey are shown for the subset of sites characterized using reverse correlation. Sites are labeled based on whether they were tested (yellow) or not tested (white). Site positions were projected to the boundary between gray and white matter before flattening of the cortical patch (black circle, approximate projection to the surface of physiologically defined boundary of PL at its widest extent in 3D; dashed contour lines are projected Horsely–Clarke AP coordinates; STS, superior temporal sulcus; similar plotting conventions as in Fig. 1*A*).

tional volumes (Jezzard and Balaban, 1995). Functional scans were acquired using an echoplanar imaging (EPI) sequence (135 time points per time series; TR, 3.25 s; TE, 30 ms; coronal phase encode, 1.25 mm slice thickness with 10% slice gap; 64 × 45 × 64 matrix; voxel size, 1.25 × 1.375 × 1.25 mm; 45 slices; coverage, −32 to +32 anteroposterior) (Op de Beeck et al., 2008). A brief saturation pulse (250 ms) was applied at the beginning of each TR to limit wrap-around artifacts from fatty tissue in the neck, and PACE (prospective acquisition correction; Siemens AG) was used to adjust for motion artifacts online. Images were presented in a block design where each block included images from a single category. Ten volumes were collected in a single block (32.5 s total duration), and each run included 12 blocks (6.5 min total duration). In a typical 2–3 h session, 10–15 runs were collected, and reliable face versus object maps were obtained within two sessions.

For each monkey, two T1-weighted EPI anatomical images (3D MPRAGE sequence, 256 × 256 × 256 voxels, 0.5 mm isotropic voxel size) were also acquired under general anesthesia (combination of ketamine and xylazine). To improve the SNR, we acquired 3–10 volumes and averaged them. The first anatomical served as a "gold standard" and was used for construction of gray/white matter boundary and pial surfaces. A second anatomical was obtained before physiological recordings once the x-ray reference frame was in place. Using this common frame, 3D x-ray based electrode coordinates could be coregistered (projected) to MRI volumes.

Analysis: MRI

The surface defined by the boundary between gray matter and white matter was created using the software Freesurfer (code publicly available at http://surfer.nmr.mgh.harvard.edu/). Briefly, volumes were motion corrected, averaged, reoriented, cropped, conformed to a 256 × 256 × 256 volume, intensity normalized, white matter normalized, skull stripped (orbits manually extracted), and renormalized in preparation for segmentation of the white matter, corpus callosum, and pons. From this segmentation, a tessellated 3D gray/white surface was generated, which was inspected for topological defects and corrected manually. The pial surface was generated by growing the gray/white boundary outward following luminance gradients. For visual display of fMRI data, the gray/white surface was inflated to expose sulci (Fig. 1*A*). Physiology data were plotted on a flattened patch of the ventral visual cortex (Figs. 1*A*, 5*B*). A patch was created by introducing cuts into the medial and dorsal aspects of the left hemisphere along with relaxation cuts. This 3D surface patch was then computationally flattened into 2D coordinates while attempting to preserve distances along the surface (near isometric flattening).

Before analyzing functional data, runs in which subjects moved more than three times were excluded. For runs that were included in analysis, volumes collected during major movements were excluded. Functional data were analyzed using a statistical parametric mapping approach where each voxel was modeled independently using FS-FAST (FreeSurfer Functional Analysis Stream) (Friston et al., 1995). Volumes were first dewarped based on the field map collected at the beginning of each session (Jezzard and Balaban, 1995). Dewarping was useful for removing geometric distortions induced by susceptibility artifacts near the ventral temporal lobe. Volumes from each session were motion corrected (Cox and Jesmanowicz, 1999) and aligned to the first volume of a session, intensity normalized, and spatially smoothed using a Gaussian kernel (FWHM, 2.5; $\sigma = 1.06$ voxels). A general linear model was used to fit parameters for the time series at each voxel in the volume and included regressors for nuisance parameters such as motion and drift. Contrast maps (e.g., faces vs objects) were computed using a group analysis that averaged across sessions.

Results

Physiological targeting of posterior face patch

We used a novel microfocal stereo x-ray system guided by fMRI maps of face versus nonface object selectivity to precisely target microelectrode recordings in two monkeys to PL with submillimeter accuracy (see Materials and Methods) (Fig. 1*A*). In our fMRI maps, PL appeared as a region in posterior IT (also referred to as TEO) bound posteriorly by the inferior occipital sulcus and followed anteriorly by the middle and anterior fMRI-determined face patches (Fig. 1*A*), consistent with the previously described fMRI pattern of face-selective regions in IT (Moeller et al., 2008; Tsao et al., 2008). Previous neurophysiological work in IT had shown strong neuronal face selectivity underlying fMRI-targeted subregions in middle and anterior IT, but it was unclear whether a subregion of posterior IT, a visual area that borders V4, would

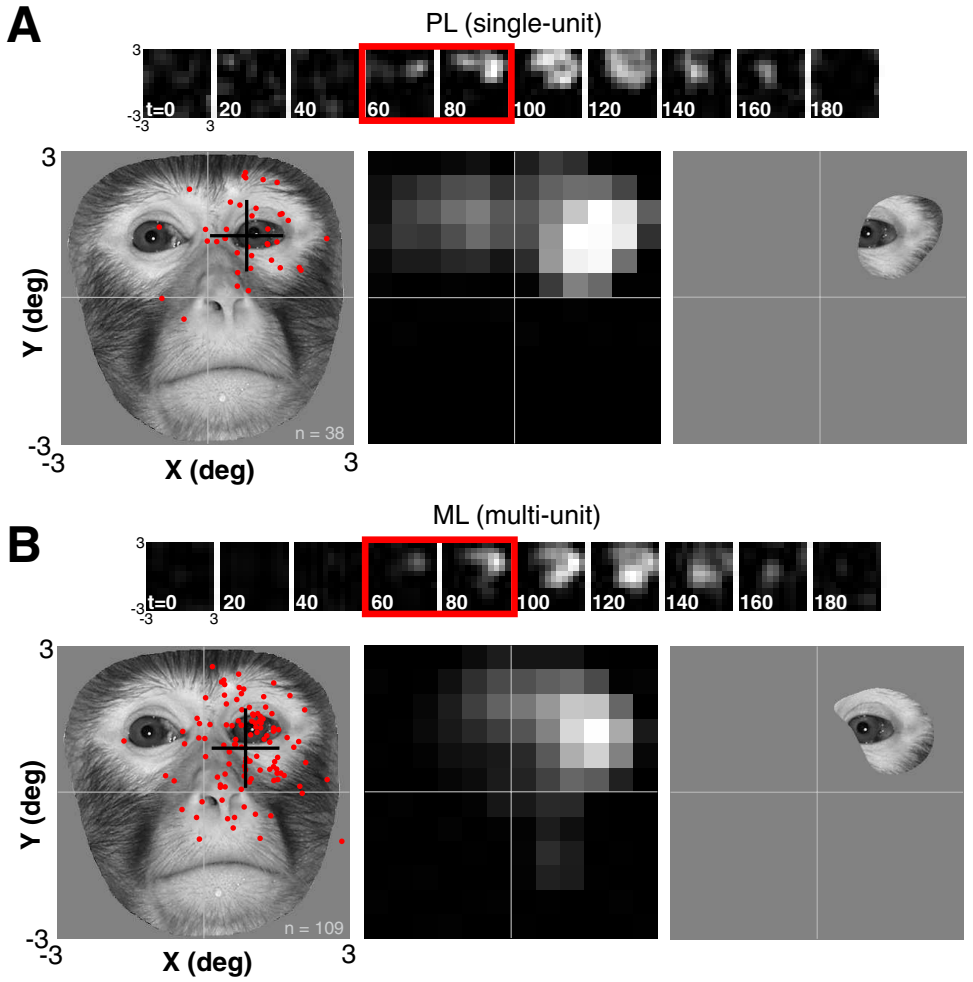

Figure 6. Reverse correlation mapping for single units and in ML. ***A***, Preferred face regions of single units recorded in PL were centered near the contralateral eye (black crosshair; compare to Fig. 3*A*; same plotting conventions used). ***B***, Preferred face regions of individual multiunit sites recorded in ML (compare to Fig. 3*A*; same plotting conventions used).

express as strong or uniform a neuronal preference for faces as more anterior regions of IT. We found, however, that when using a standard definition for face cells (response to faces greater than twice the response to nonfaces), PL contained a similar fraction of such sites as the middle lateral (ML) and anterior [anterior medial (AM) and anterior lateral (AL)] face patches [PL, 83%; ML, 75%; AM/AL, 88% of sites with FSI $>$ 1/3; n = 150, 420, 82 sites, respectively] and demonstrated similarly high selectivity (mean FSI, PL, 0.65; ML, 0.56; AM/AL, 0.74). Although we primarily recorded multiunit activity, the face selectivity distribution of multiunits was comparable to that for single units (Fig. 1*C*). Sites in PL responded, on average, much more strongly (Fig. 1*B*) and with shorter latency by ~12 ms (Fig. 1*D*, left, middle) to images of faces than to images of bodies, objects, and places (median, 80 vs 92, 92, and 94 ms, respectively). Response latencies of neurons in PL were ~5 ms shorter than in the middle and anterior patches, suggesting that PL is an earlier stage of visual processing (median latency, PL, 74; ML, 79; AM/AL, 80 ms; $p <$ 0.01). The relative preference for face images in PL peaked in the first 40 ms of the response (Fig. 1*D*, right), and we chose to focus on this first wave of activity to test what features engage face selectivity in IT.

Responses to face fragments in PL

We used a subspace reverse correlation technique (Fig. 2*A*) (see Materials and Methods) to systematically map which parts of the face image drove each PL recording site. Our goal was not to build models for each neuron as done in some physiological studies, but rather to obtain the spatial weighting across a typical face image similar to previous psychophysical work (Gosselin and Schyns, 2001). Briefly, this method rapidly exposes small, randomly selected regions on face images (47 ms per image) and provides an unbiased estimate of each region's importance in driving the earliest neuronal spikes (60–100 ms time window following image onset) (Gosselin and Schyns, 2001). For example, the site in Figure 2*A* showed strong modulation during presentation of consecutive face fragments. Visual inspection reveals that epochs where the interior portion of the eye was present drove responses to similar levels as for the whole face (Fig. 2*A*, red highlighted region), and this observation was consistent with the linear weight map derived using reverse correlation between spikes and images (Fig. 2*B*).

Almost all sites across PL were modulated by specific subregions of the face (median, 15 bins $>$ 1.5 SDs above baseline; 144 total bins; corresponds to 3.75 deg^2; only one site yielded no significant bins; n = 111 sites). And in an overwhelming majority of sites, the quadrant in the upper contralateral field that contained the eye was highly important, while the lower field that contained the mouth had little influence (108 of 111 sites had a preferred region centered in the upper, contralateral quadrant) (Fig. 3*A*, red dots; white dot indicates center of preferred region for the example site from Fig. 2; for the preferred regions of all

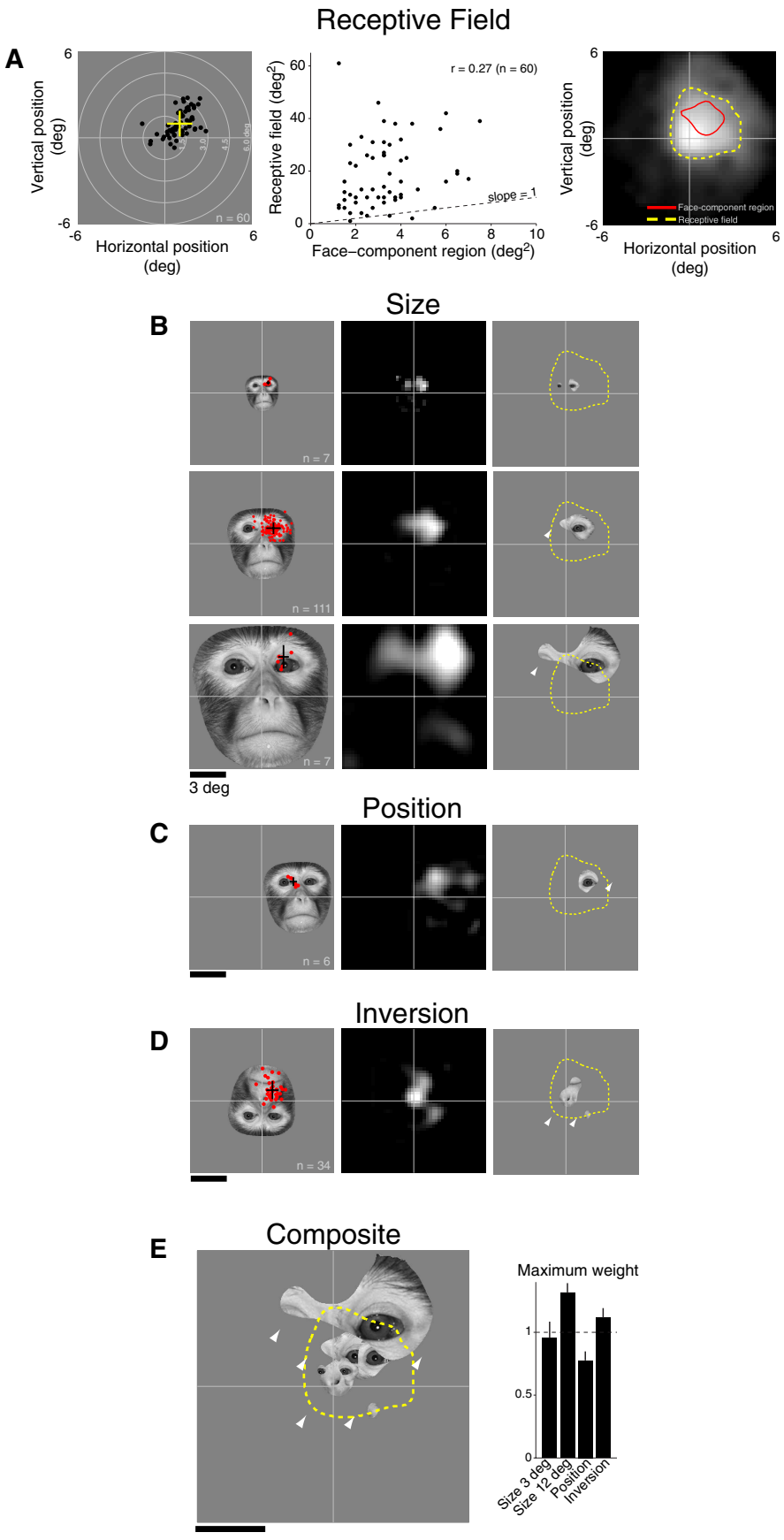

Figure 7. Size and position tolerance of eye selectivity in PL. ***A***, Left, Retinal RFs in PL (black dots reflect receptive field centers of individual sites; crosshair reflects mean and SD of centers across the population). Middle, The size of RFs was only weakly predictive of the size of the preferred face components of individual sites, and retinal RFs tended to be larger than preferred face components. Right, The population retinal RF (yellow outline) was larger in size than the population-averaged face-component region (red). ***B***, The eye region was preferred (*Figure legend continues.*)

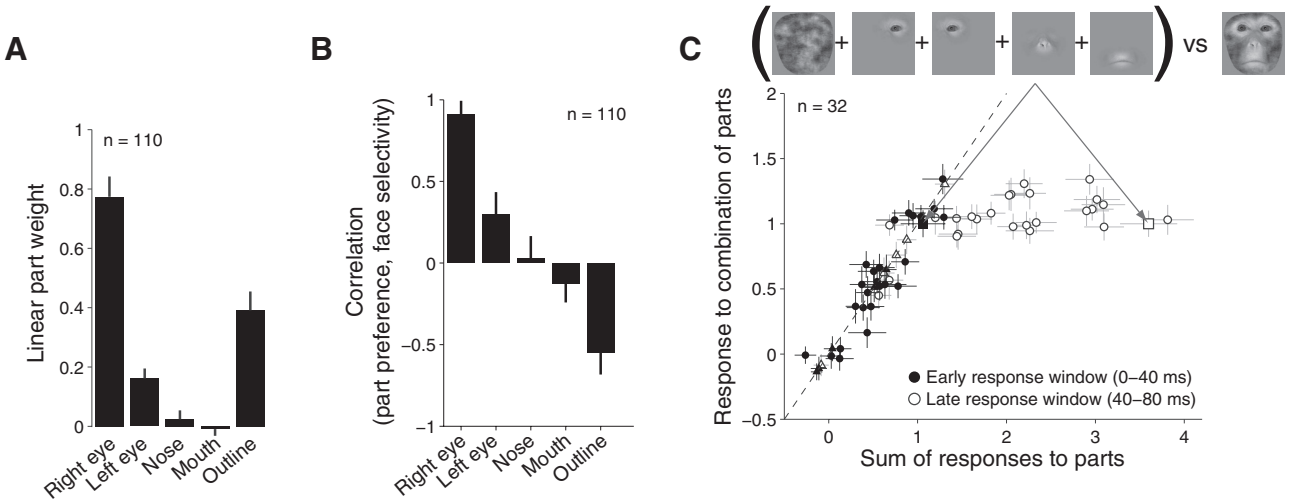

Figure 8. Linear contribution of face outline and contralateral eye. ***A***, The linear contribution of each human-segmented face part (right eye, left eye, nose, or mouth) when testing all possible combinations ($2^4 = 16$ total images) in the context of a face outline. ***B***, Correlation of part preference (i.e., $w_{part}/\sum|w_j|$) with face selectivity index. ***C***, The median response of the population to all possible part combinations where the presence of the outline was also varied ($2^5 - 1 = 31$ total images) plotted against the sum of the responses to the parts (vertical bars, SD of responses across the sites tested; horizontal bars, SD of the sum of parts responses across the sites tested; triangles, single-part responses; square symbols, whole-face response vs the sum of the responses to the five parts). Prior to averaging across sites, the responses for each site were rescaled by the average response to the whole face.

individual sites, see Fig. 4). This overall preference for the contralateral eye region could not be simply explained by the higher local contrast of the eye compared to other parts of the face, as the preference for the contralateral eye region persisted in control faces that manipulated local contrast and spatial frequency to be more similar across the image but maintained the form of the eye (Fig. 3*B*, left, middle). The eye region was also preferred in reversed-contrast faces, suggesting that the dark appearance of the eye was not simply conferring an advantage over the remainder of the face, which was lighter in appearance (Fig. 3*B*, right); however, these contrast manipulations did reduce the overall response weighting compared to the original face image (Fig. 3*B*, far right). There was little evidence that monkeys were preferentially looking toward the eye, as fixation distributions were maintained within <0.5° of the central fixation spot (90% confidence interval, 0.55°) and had little radial bias in any direction, similar to the pattern of fixations on our general screen set (90% confidence interval, 0.46°) (Fig. 5*A*). The observed consistency of neuronal responses could not be explained by oversampling sites in a particular region or a particular monkey as sites tested with reverse correlation mapping were a large fraction of all PL sites recorded ($n_{revcorr}$ = 111/150 total) and were spatially sampled throughout the PL patch in both monkeys (Fig. 5*B*). Finally, one possibility is that recording multiunit activity averages across interesting variation at the single neuron level, but we also found that the eye region was generally preferred in the single units that we sampled (Fig. 6*A*). Reverse correlation maps in the middle face patch also revealed a general preference toward the contralateral eye region, although this preference was slightly less consistent than in the posterior face patch (87 of 109 sites had a preferred region centered in the upper, contralateral quadrant) (Fig. 6*B*).

To test the importance of the eye region in driving face selectivity as measured using the FSI (Tsao et al., 2006; Freiwald and Tsao, 2010), we computed a measure of eye preference (the fraction of face-component weights in a 1.5 × 1.5° window over the eye region) (Fig. 3*C*, left) and found that it moderately predicted face versus nonface object selectivity across sites in the posterior face patch (r = 0.43, n = 111 sites) (Fig. 3*C*, right). In contrast, preference for other regions of the face (computed as the fraction of weights in five nonoverlapping 1.5 × 1.5° windows that excluded the contralateral eye) was not as strongly correlated with face selectivity (Fig. 3*C*, middle). Thus, the strength of tuning to the eye region predicted the degree of face selectivity as previously defined operationally on general image sets (Tsao et al., 2006; Freiwald and Tsao, 2010), even for a sample of sites where the majority already passed the minimum operational criterion for face selectivity (FSI > 1/3, or response to faces greater than twice the response to nonfaces).

Retinal receptive fields in PL

When measured by flashing a 3° face at random positions on a polar grid (6° radius), PL retinal receptive fields were found to be systematically biased toward the upper contralateral quadrant (mean azimuth, 1.04°; mean elevation, 0.99°) (Fig. 7*A*, left) consistent with previous work showing topographical organization of upper versus lower field preferences in posterior IT or area TEO (Boussaoud et al., 1991; Brewer et al., 2002). This suggests that recovered face components (Figs. 3, 4) may simply reflect the retinal receptive fields of sites in PL. However, retinal receptive fields were nearly six times larger (area at half-height) than preferred face regions measured using reverse correlation (Fig. 7*A*, right), and we found a mild correlation between retinal receptive

←

(*Figure legend continued.*) across size changes of two octaves. ***C***, The left eye region was preferred over the right eye region when the face was shifted horizontally into the contralateral field. ***D***, The centers of reverse correlation maps did not shift completely to the lower field even though the eye region was in the lower field in inverted faces. Population-averaged weight maps in ***B–D*** show only the positive weights (negative weights were minimal) and are normalized by the peak response to the standard image condition (6° upright face centered on fixation; see ***E*** for comparison of peak magnitudes across conditions). Recovered image fragments are based on 50% contours of excitatory weight maps. ***E***, Composite of 50% response regions in ***B–D***. Pupil centers where no fragments were recovered fell outside the retinal aperture (white arrows). The magnitude of the maximum weights for recovered fragments within the retinal aperture was similar to the magnitude of weights for the original face image (dashed line; right). Scale bars, 3°.

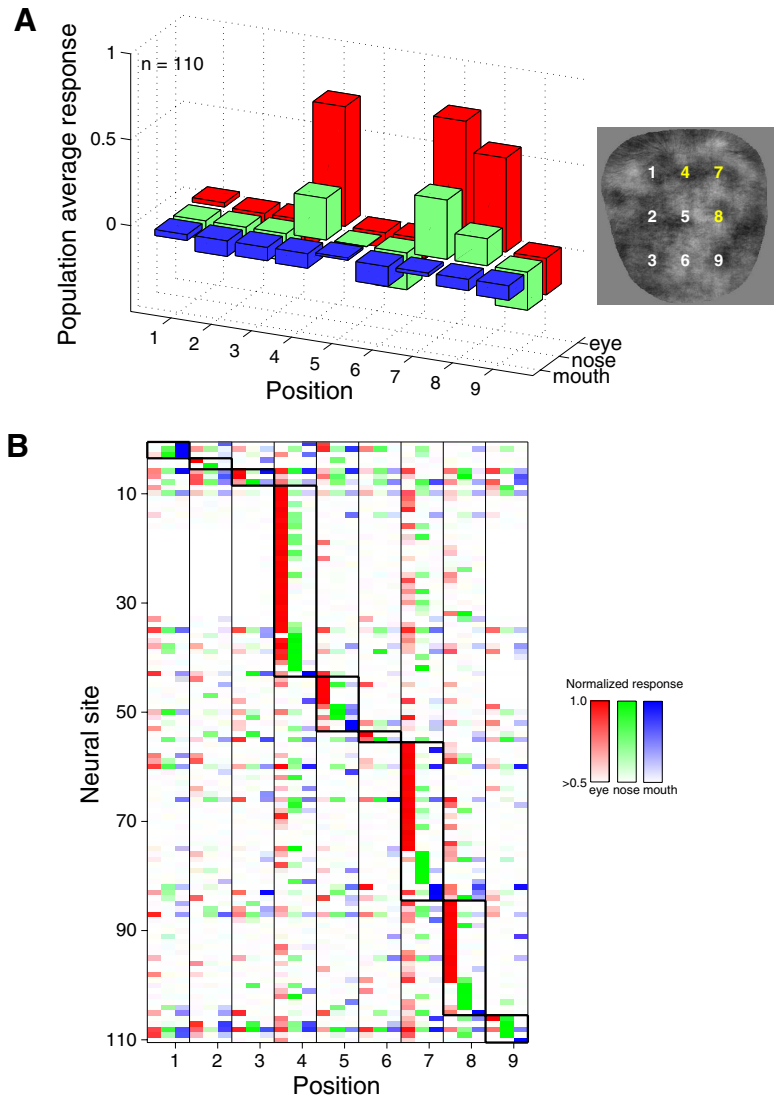

Figure 9. Effect of retinal position on responses to face parts. ***A***, Population-average responses to the face parts (eye, nose, and mouth) tested at nine possible positions on the face outline (the response to the outline alone was subtracted, and responses were normalized by the median outline-subtracted response to the whole face). ***B***, Bottom, Color histogram showing the responses of individual sites to the eye, nose, and mouth presented at nine different positions. Sites are ordered by their position preference (black boxes) followed by their feature preference at the preferred position (i.e., eye-selective sites appear first followed by nose- and then mouth-selective sites). Responses are thresholded at half-maximum (see color legend).

field sizes and the size of preferred face components ($r = 0.27$; $p > 0.01$; $n = 60$) (Fig. 7*A*, middle), suggesting that retinal receptive fields could not fully account for selectivity to the eye region. Rather, PL neurons behaved as if an upper, contralateral retinal aperture constrained the visual field region over which they expressed a preference for eye-like features. Within this aperture, the highest weights in reverse correlation maps shifted to track the position (0.8, 1.6, and 3.2° shifts) and the size of the eye region in faces that were one octave larger or one octave smaller than the original size tested (Fig. 7B). Furthermore, although the right eye was a strong driver of initial PL responses when the face image was at the center of gaze (Fig. 7*B*, middle), the left eye became an important driver of the response when the face was made small enough (3°) such that the whole face fell within the retinal receptive field (Fig. 7*B*, top), or when the face was shifted horizontally so that the left eye was now in the contralateral retinal field (Fig. 7*C*). This contralateral shift manipulation, which moved the right eye region to the edge of the retinal receptive field, reduced the previously strong response to that eye region (Fig. 7*C*). Finally, inversion of the face, which moved the eyes to the lower hemifield, diminished the response to the eye region (Fig. 7*D*). Curiously, inversion led to a response near the center of gaze to an upturned nose where, purely for the sake of speculation, the nostril could be viewed as a darkened, round region much like an eye in the upper field position where an eye might be expected (Fig. 7*E*). Importantly, the magnitude of response weights remained strong for these manipulations of the original face image even though the exact features in the retinal aperture were changing in their size, position, and spatial frequency (Fig. 7*E*, right). Together, these observations suggest a qualitative first-order response model in which PL neurons are tuned to eye-like features within an upper, contralateral field retinal aperture (mean retinal location, 1.04° azimuthal, 0.99° elevation; mean size, 4.6° in diameter) (Fig. 7*E*).

Responses to the combinations of face parts

While the reverse correlation approach provided strong clues as to the key image features driving early PL responses (Figs. 3, 4), we were concerned that it only tested a limited subspace of face features, namely, local fragments limited to the size of our Gaussian window (~1/16 of the face) (Fig. 2*A*); that is, neural responses may depend on larger-scale features (e.g., face outline) or combinations of parts (e.g., both eyes) that were not explicitly tested in our experiments. Closer inspection of the preferred image fragments of individual PL sites reveals that when these regions were not constrained to the eye, they often contained portions of the external outline and not the nose or mouth (Fig. 4). Thus, to test larger face regions including the whole outline, we presented all possible combinations of the face parts (ipsilateral eye, contralateral eye, nose, and mouth) in the context of the outline ($2^4 = 16$ total images) using standard rapid serial visual presentation (100 ms on, 100 ms off) (Hung et al., 2005; De Baene et al., 2007). We computed the linear weights of the five tested face parts in this combinatorial image set. We found that the contralateral eye was the most strongly weighted part (Fig. 8*A*), a result consistent with what we found from rapid reverse correlation mapping with a completely different set of visual stimuli. A novel observation, however, was that the face outline, which was not directly tested in reverse correlation mapping, was also important in driving responses. This contribution was not as strong as the contribution of the contralateral eye, and PL sites preferring the outline tended to be on the lower end of the distribution of face selectivity, unlike sites preferring the eye (units in the upper quartile of eye preference, mean FSI, 0.97; $n = 28$; units

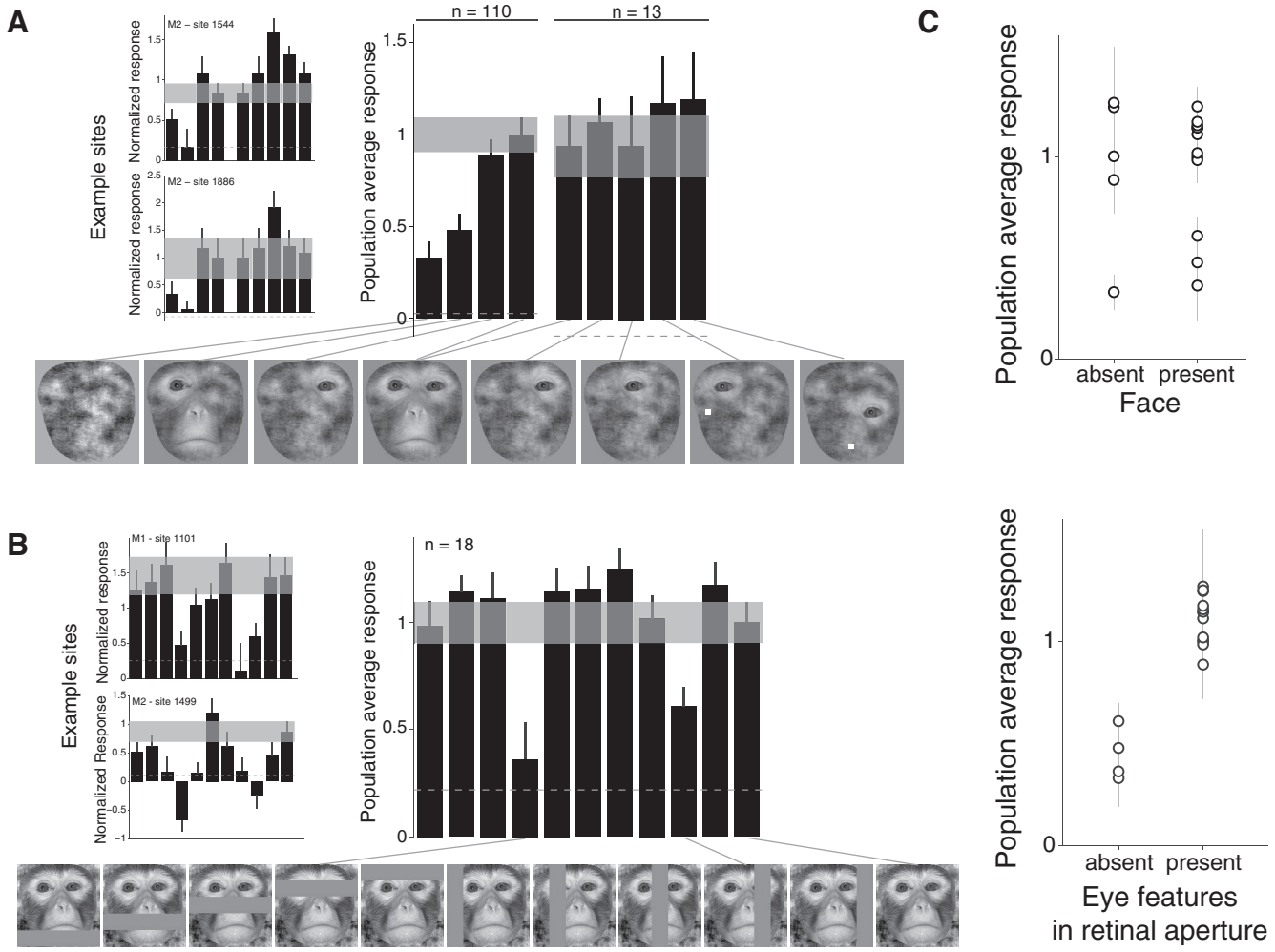

Figure 10. Images predicted to drive early responses in PL. ***A***, Left, Responses to images containing some or all of the face parts in the context of the face outline. Right, The outline-relative position of the eye could be varied across a wide range and still drive early responses in PL sites. The white square indicates the fixation point when not on the center of the image. Insets, Responses of two example sites to the same images. ***B***, Reduced early responses for faces were found when a horizontal or vertical bar occluded the contralateral eye. Insets, Responses of two example sites to horizontally or vertically occluded faces. ***C***, Top, PL responses were poorly predicted by whether a face was present or absent (face considered present if at least half of the face is visible and absent otherwise). Bottom, PL responses were relatively well predicted by whether eye-like features were in the upper contralateral visual field. Raw average of individual site responses shown in ***A*** and ***B*** were rescaled by the median response to the whole face. Dashed lines indicate the median response to nonface objects.

in the upper quartile of outline preference, mean FSI, 0.59; $n = 28$; all units, mean FSI, 0.76; $n = 110$; part selectivity defined as $w_{\text{part}}/\Sigma|w_i|$) (Fig. 8*B*). Regardless, both the eye and the outline were needed to fully modulate early PL responses.

We next asked whether the face parts (e.g., the eye and outline) are combined linearly or demonstrate a nonlinear interaction by varying the presence or absence of all parts of the face (right eye, left eye, nose, mouth, and outline; $2^5 - 1 = 31$ possible images). We found that early-phase responses to combinations of parts (two, three, four, and five parts) were well matched by a linear sum of responses to the individual parts (median noise-adjusted correlation, 0.95; $n = 32$ PL sites). Responses fell along the unity line arguing against sublinear (averaging and normalization) or superlinear (conjunction sensitivity) integration of these parts (Fig. 8*C*). On the other hand, late-phase responses (41–80 ms after response onset) were less well fit by a linear model (noise-adjusted correlation, 0.59; $n = 32$ PL sites) and demonstrated sublinear integration of parts (sum of responses to the parts exceeded the response to the whole) similar to findings in the middle face patch (Freiwald et al., 2009) (Fig. 8*C*, open square; note the large deviation from the unity line of the sum of the late responses to all five parts vs the late response to the whole face).

Control for the effect of retinal position

Our results thus far suggest that three factors, eye-like image features, boundary features (e.g., face outline), and retinal position, are dominant factors in driving PL early responses. To explicitly control for the influence of retinal position, we presented the face parts (i.e., eye, nose, and mouth) at nine positions across the face outline. Only when the eye was placed in the contralateral field were responses maximized. Placing the nose in the upper, contralateral field gave a weaker response than the eye, and the mouth gave almost no response when tested in the upper contralateral field (Fig. 9*A*). Across positions in the contralateral visual field, the eye was consistently preferred by a majority of sites (part eliciting maximal response, 65% eye, 25% nose, and 10% mouth; $n = 110$ sites), suggesting the importance of eye-like image features independent of retinal position in eliciting maximal responses (Fig. 9*B*).

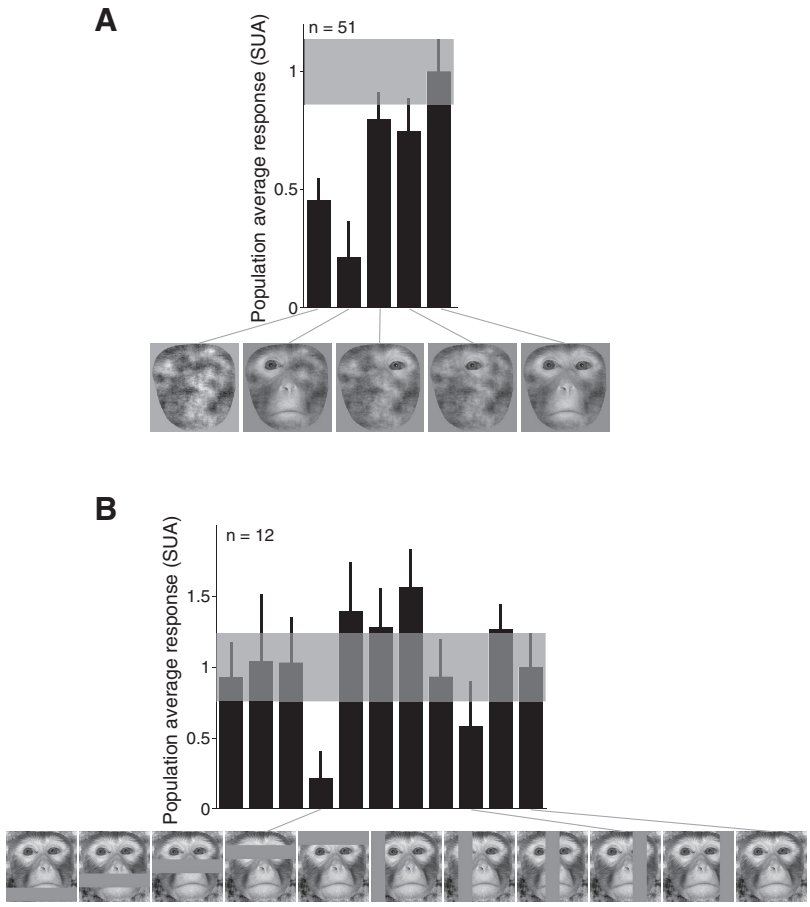

Figure 11. Responses of single units when the eye region was absent or occluded. ***A***, Single-unit responses were highest when an eye was present in the upper, contralateral field and were strongly reduced when the contralateral eye was absent, even though all other face parts were present. ***B***, Single-unit responses were strongly reduced when both eyes were occluded and weakly reduced when the contralateral eye was occluded.

Images predicted to drive early responses in PL

Given the importance of the eye and its retinal position in the context of a curved boundary such as the face outline, we make some key predictions that diverge qualitatively from the common intuition that face cells are whole-face detectors. First, our qualitative model predicts that normal face images lacking only an eye in the PL retinal aperture should not produce strong responses. We found this to be the case: PL neurons gave a very weak response to this image (Fig. 10*A*, left), even though it evokes a strong face percept. Similarly, under more naturalistic conditions where a face is always present behind an occluder, occluding the contralateral eye region strongly decreased the initial response, while occluding other regions of the face did not significantly affect the early response (Fig. 10*B*). Second, our qualitative model predicts that simply presenting a nonface image that contains an eye in an outline should be sufficient to strongly drive PL responses. Indeed, this prediction was born out: responses to this reduced, nonface image nearly matched the response to the entire face (Fig. 10*A*, left). Extending this prediction further, our qualitative model suggests that the exact location of the eye in the outline should not matter, as long as it is on a curved boundary within the retinal aperture. Indeed, we found that "cyclops-like" eye arrangements without any other face features drove PL neurons as well as full-face images (Fig. 10*A*, right), and placing the eye on the opposite side of the outline (or lowering the eye to the middle of the outline) elicited a strong response (when compensated by an offset in fixation; Fig. 10*A*, right). Thus, whether the eye was in the right, middle, or left outline-centered position, similar PL responses were obtained to all of these nonface images, presumably because all of these images shared the property that eye-like features and some portion of a curved outline were within the retinal aperture. Together, these tests are consistent with our qualitative response model. More deeply, this new understanding allowed us to create stimuli that reveal a complete disconnect between neuronal responses and the human notion of what is and what is not a face (Fig. 10*C*, top). Rather, the mean population response in PL signaled whether eye-like features are physically present or absent within a fixed retinal aperture (Fig. 10*C*, bottom).

Comparison of single-unit, multiunit, and population-average responses

Previous work in fMRI-determined primate face patches has recorded single units (Tsao et al., 2006), while we mainly focused on multiunits recorded as part of our mapping studies. For the images that we tested, we found similar properties between multiunits and the single units we sampled. Single units demonstrated face selectivity comparable to that of multiunits (Fig. 1*C*), preferred the contralateral eye in reverse correlation maps (Fig. 6*A*), and were sensitive to removal or occlusion of the eye but not other face parts (Fig. 11). There was good agreement in rank-order image selectivity between multiunits and single units recorded at the same site (screen set, $r = 0.99$, $n_{images} = 40$, $n_{sites} = 46$; part combinations/positions, $r = 0.95$, $n_{images} = 39$, $n_{sites} = 19$; correlations were noise adjusted). Furthermore, multiunit sites were strongly correlated with the population average, justifying the use of population averages in our main analysis (screen set, $r = 0.78$, $n_{images} = 40$, $n_{sites} = 129$; part combinations/positions, $r = 0.80$, $n_{images} = 39$, $n_{sites} = 76$; occlusion, $r = 0.81$, $n_{images} = 11$, $n_{sites} = 14$). This functional clustering suggests that the properties we have reported generalize across PL.

Discussion

Here, we have demonstrated that eye-like features in the upper, contralateral visual field are a key building block of the early responses of PL face cells. Selectivity for eye-like features was a predictor of standard measures of face selectivity (Figs. 3*C*, 8*B*), and a single eye placed at a wide range of positions inside a curved boundary such as the face outline was sufficient to produce responses as strong as those to normal faces (Fig. 10*A*). Given that the face outline proved critical in driving the highest responses in PL, this could be viewed as evidence for holistic face processing (Tanaka and Farah, 1993), but we note that the eye and outline could strongly drive responses even when presented alone (Fig. 8*A*), summed linearly (Fig. 8*C*), did not have to be in a natural

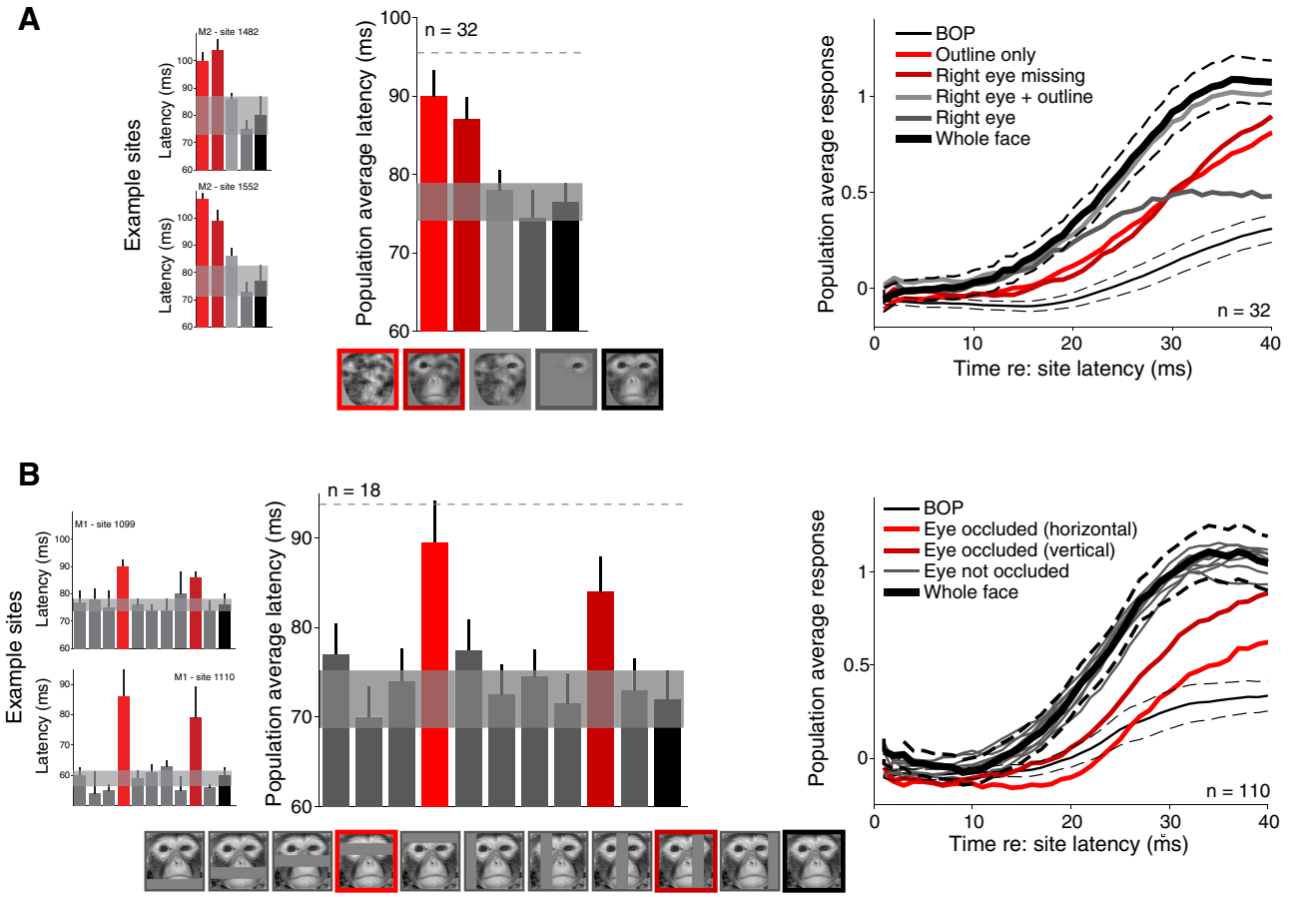

Figure 12. Response latency when the eye region was absent or occluded. ***A***, Simply presenting the right eye (dark gray) but not the outline (light red) led to a response latency as short as for the whole face (black). Removing the right eye delayed responses even when all of the other face parts were present (dark red). Insets, Response latencies in two example sites for the same set of images. ***B***, Latency increased when occluding the contralateral eye region using a horizontal or vertical bar (red) but not when occluding other regions of the face (gray). Insets, Response latencies of two example sites to horizontally and vertically occluded images. ***A***, ***B***, Right, Population-averaged PSTHs shown rescaled by the median response to the whole face at 100 ms. Individual site PSTHs were aligned by their minimum latency to faces, bodies, objects, or places before averaging.

configuration (Fig. 10*A*), and tended to drive different sites (correlation between linear weight of outline and contralateral eye, -0.75; $p < 0.01$; $n = 110$), suggesting independent processing of these features. Put another way, one possible qualitative view of PL responses is that PL receives independent inputs about eye-like features and a boundary curve (each within a retinal aperture), and then simply sums those inputs. The specific details of these inputs and their dependence on the exact form, contrast, position, and size of eye-like image features or outline image features remain to be determined. We have only used the descriptors "eye-like" and "outline" as a proxy for the true image preferences of PL cells, as the goal of this study was not to build full descriptions of PL receptive fields, but to identify the regions of the face that trigger face-cell responses. Our results take a first step toward this goal by showing that, instead of treating these cells as true whole-face detectors, a parts-based view may be more appropriate. How the remaining face parts besides the eye and outline are represented is an important open question. Processing of different face regions may be performed in patches or pathways besides PL (e.g., superior temporal sulcus face patches), may be reflected in later responses within PL (Fig. 8*C*, open triangles), or may reside completely outside the currently defined "face network" (Tsao and Livingstone, 2008).

All previous work on face-selective cells has been in the middle (Tsao et al., 2006; Freiwald et al., 2009; Freiwald and Tsao, 2010) or anterior (Freiwald and Tsao, 2010) face patches, while we report here on cells in the posterior face patch—a cortical subregion that is likely to be an earlier stage of face processing. It is worth noting, however, that the posterior patch exhibited comparable face selectivity (on standard measures) to the middle and anterior patches. When we measured reverse correlation maps in the ML, we also found that early responses tended to be driven by the eye region (Fig. 6*B*), which is compatible with the notion that ML may receive feedforward input from PL and with previous work showing that these two areas are functionally connected (Moeller et al., 2008). Indeed, a previous study found that ML neurons are primarily sensitive to contrast polarity in the region of the eyes (Ohayon et al., 2012). In the fusiform face area in humans, the eyes alone do not drive as strong a BOLD signal as the face without the eyes (Tong et al., 2000), but this may be because the BOLD signal integrates over much longer timescales (seconds) than the early physiological window (milliseconds) in which we revealed the importance of eye-like features for driving responses in monkey PL neurons. A possible hypothesis is that area PL in monkeys corresponds more closely to the occipital face area (OFA) in humans, as the OFA is considered to be an earlier stage of face processing than the fusiform face area (Gauthier et al., 2000). Indeed, the contralateral eye is sufficient to drive early evoked potentials (N170) and MEG responses in the OFA in

humans (Schyns et al., 2003; Smith et al., 2004, 2009) in parallel to the findings presented here for PL in monkeys.

Previous work using analysis windows (>100 ms) that favored late-phase responses showed sensitivity to all parts of the face in the context of the outline, not just the eye region (Freiwald et al., 2009). Here, we sought to gain insight into the first steps of face processing and focused on the initial response transient (first 40 ms) under the logic that this should be the most "feedforward" component of the response (Hung et al., 2005) (and thus the easiest to explain) (Brincat and Connor, 2006) and motivated by the observation that standard face selectivity peaks within this time window (Fig. 1*D*). Much like in previous work (Freiwald et al., 2009), we found that face parts besides the contralateral eye and outline elicited responses in a later time window (40–80 ms) (Fig. 8*C*, open triangles; late responses to single parts were >0.5 of the response to the whole face, except for the mouth), and the sum of the responses to the parts exceeded the response to the whole (i.e., sublinear summation, unlike in early-phase responses) (Fig. 8*C*). The exact nature of nonlinear processing in the late phase of PL responses will be the subject of future work and may provide insights into the dynamics of face inference.

Our results in the early response window argue against models for face detection that initially rely on whole-face templates (Tsao and Livingstone, 2008), nonlinear interactions between parts, or object-centered configuration of parts. Rather, we present the case that bottom-up face processing is relatively local (Fig. 3), linear in a simplified feature space (Fig. 8*C*), and retinally centered, as it was critical for the eye to remain in the same retinal position even when the outline was shifted (Figs. 7*C*, 10*A*). Though limited to a retinal aperture, eye selectivity in PL exhibited some position and size tolerance (Figs. 7*E*, 9) and could not be simply explained by low-level factors such as contrast and spatial frequency (Fig. 3*B*), but we do not rule out a model that includes these factors in combination. These field size, selectivity, and tolerance properties are consistent with the intermediate position of PL in the ventral visual hierarchy (Op de Beeck and Vogels, 2000; Brincat and Connor, 2004; Rust and DiCarlo, 2010) and provide constraints for models of early face selectivity. We caution, however, that these properties do not yet provide a quantitative (i.e., image computable) model of PL responses. Future work will aim at exactly specifying the image operators that explain these neuronal responses, and our data suggest that these operators will have an "intermediate complexity" predicted by computational work on face detection (Ullman et al., 2002).

Although our results suggest that early face detection does not rely on a true matched filter for faces, there is likely a high correlation in the natural world between the presence of a face and the presence of the eye and boundary features preferred by PL neurons (Ullman et al., 2002). Notably, the mean center of PL retinal preference regions (mean azimuth, 1.04°; elevation, 0.99°) corresponds to the expected location of eye features in upright, real-world monkey faces viewed at a distance of ~1 m (assuming that gaze is centered on the nose, as suggested by saccade landing distributions for faces) (Hsiao and Cottrell, 2008). Specialization to the statistics of faces under real-world viewing conditions may improve speed for commonly encountered images while sacrificing accuracy for less encountered images such as inverted faces (Lewis and Edmonds, 2003; Sekuler et al., 2004). That detection of upright faces is relatively rapid compared to other object classes has been documented in both neural (Bell et al., 2009) and psychophysical studies (Crouzet et al., 2010; Hershler et al., 2010) and was also evident in our data (Fig. 1*D*). And we suggest that eye-like features are contributing to this rapid response, as presenting the eye alone was sufficient to elicit response latencies similar to those for the whole face (Fig. 12*A*). Removing or occluding the features in the eye region while maintaining the presence of the face outline or other face parts slowed response latencies to upright faces by ~10 ms (Fig. 12*A*,*B*). This delay is consistent with increased reaction times in psychophysical tasks when the eyes are occluded (Fraser et al., 1990; Lewis and Edmonds, 2005). Remarkably, occluding any other region of the face had no effect on human reaction times (Fraser et al., 1990; Lewis and Edmonds, 2005) nor on PL neuronal response latencies (Fig. 12*B*). The importance of rapid detection of the eyes may go beyond face detection, as the eyes are the primary target of saccades following the central landing saccade on the nose (Walker-Smith et al., 1977; Henderson et al., 2005; Hsiao and Cottrell, 2008) and contain information important for judging gaze, identity, gender, and emotion (Schyns et al., 2002, 2007; Itier and Batty, 2009). During development in children, eye processing appears to precede maturation of full-face processing (Bentin et al., 1996; Taylor et al., 2001). Finally, the importance of the eyes is consistent with many computer vision systems, such as the Viola and Jones (2001) face detector, that rely on features in the eye region for performing frontal face detection tasks, and biology may have converged on a similar solution that is both simple and effective.

References

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. Medline

Bentin S, Allison T, Puce A, Perez E, McCarthy G (1996) Electrophysiological studies of face perception in humans. J Cogn Neurosci 8:551–565. CrossRef Medline

Boussaoud D, Desimone R, Ungerleider LG (1991) Visual topography of area TEO in the macaque. J Comp Neurol 306:554–575. CrossRef Medline

Brewer AA, Press WA, Logothetis NK, Wandell BA (2002) Visual areas in macaque cortex measured using functional magnetic resonance imaging. J Neurosci 22:10416–10426. Medline

Brincat SL, Connor CE (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat Neurosci 7:880–886. CrossRef Medline

Brincat SL, Connor CE (2006) Dynamic shape synthesis in posterior inferotemporal cortex. Neuron 49:17–24. CrossRef Medline

Buades A, Le TM, Morel JM, Vese LA (2010) Fast cartoon + texture image filters. IEEE Trans Image Process 19:1978–1986. CrossRef Medline

Cox DD, Papanastassiou AM, Oreper D, Andken BB, Dicarlo JJ (2008) High-resolution three-dimensional microelectrode brain mapping using stereo microfocal X-ray imaging. J Neurophysiol 100:2966–2976. CrossRef Medline

Cox RW, Jesmanowicz A (1999) Real-time 3D image registration for functional MRI. Magn Reson Med 42:1014–1018. CrossRef Medline

Crouzet SM, Kirchner H, Thorpe SJ (2010) Fast saccades toward faces: face detection in just 100 ms. J Vis 10:16.11–17. CrossRef Medline

De Baene W, Premereur E, Vogels R (2007) Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. J of Neurophysiology 97:2900–2916. 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

DiCarlo JJ, Johnson KO (1999) Velocity invariance of receptive field structure in somatosensory cortical area 3b of the alert monkey. J Neurosci 19:401–419. Medline

DiCarlo JJ, Maunsell JH (2005) Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. J Neurophysiol 93:2974–2986. CrossRef Medline

Fraser IH, Craig GL, Parker DM (1990) Reaction time measures of feature saliency in schematic faces. Perception 19:661–673. CrossRef Medline

Freiwald WA, Tsao DY (2010) Functional compartmentalization and view-

point generalization within the macaque face-processing system. Science 330:845–851. CrossRef Medline
Freiwald W, Tsao D, Livingstone M (2009) A face feature space in the macaque temporal lobe. Nat Neurosci 12:1187–1196. CrossRef Medline
Friedman HS, Priebe CE (1998) Estimating stimulus response latency. J Neurosci Methods 83:185–194. CrossRef Medline
Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RS (1995) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210.
Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform "face area" is part of a network that processes faces at the individual level. J Cogn Neurosci 12:495–504. CrossRef Medline
Gosselin F, Schyns PG (2001) Bubbles: a technique to reveal the use of information in recognition tasks. Vision Res 41:2261–2271. CrossRef Medline
Gross CG, Rocha-Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. J Neurophysiol 35:96–111. Medline
Henderson JM, Williams CC, Falk RJ (2005) Eye movements are functional during face learning. Mem Cognit 33:98–106. CrossRef Medline
Hershler O, Golan T, Bentin S, Hochstein S (2010) The wide window of face detection. J Vis 10:21. CrossRef Medline
Hsiao JH, Cottrell G (2008) Two fixations suffice in face recognition. Psychol Sci 19:998–1006. CrossRef Medline
Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. Science 310:863–866. CrossRef Medline
Itier RJ, Batty M (2009) Neural bases of eye and gaze processing: the core of social cognition. Neurosci Biobehav Rev 33:843–863. CrossRef Medline
Jezzard P, Balaban RS (1995) Correction for geometric distortion in echo planar images from B0 field variations. Magn Reson Med 34:65–73. CrossRef Medline
Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol 71:856–867. Medline
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
Lewis MB, Edmonds AJ (2003) Face detection: mapping human performance. Perception 32:903–920. CrossRef Medline
Lewis MB, Edmonds AJ (2005) Searching for faces in scrambled scenes. Visual Cogn 12:1309–1336. CrossRef
Loy G, Zelinsky A (2003) Fast radial symmetry for detecting points of interest. IEEE Trans Pattern Anal Mach Intell 25:959–973. CrossRef
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
Ohayon S, Freiwald WA, Tsao DY (2012) What makes a cell face selective? The importance of contrast. Neuron 74:567–581. CrossRef Medline
Op de Beeck H, Vogels R (2000) Spatial sensitivity of macaque inferior temporal neurons. J Comp Neurol 426:505–518. CrossRef Medline
Op de Beeck HP, Deutsch JA, Vanduffel W, Kanwisher NG, DiCarlo JJ (2008) A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. Cereb Cortex 18:1676–1694. 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
Portilla J, Simoncelli EP (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. Int J Comput Vision 40:49–70. CrossRef
Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput 16:1661–1687. CrossRef Medline
Rolls ET (1984) Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. Hum Neurobiol 3:209–222. Medline
Rust NC, DiCarlo JJ (2010) Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J Neurosci 30:12978–12995. CrossRef Medline
Schyns PG, Bonnar L, Gosselin F (2002) Show me the features! Understanding recognition from the use of visual information. Psychol Sci 13:402–409. CrossRef Medline
Schyns PG, Jentzsch I, Johnson M, Schweinberger SR, Gosselin F (2003) A principled method for determining the functionality of brain responses. Neuroreport 14:1665–1669. CrossRef Medline
Schyns PG, Petro LS, Smith ML (2007) Dynamics of visual information integration in the brain for categorizing facial expressions. Curr Biol 17:1580–1585. CrossRef Medline
Sekuler AB, Gaspar CM, Gold JM, Bennett PJ (2004) Inversion leads to quantitative, not qualitative, changes in face processing. Curr Biol 14:391–396. CrossRef Medline
Smith ML, Gosselin F, Schyns PG (2004) Receptive fields for flexible face categorizations. Psychol Sci 15:753–761. CrossRef Medline
Smith ML, Fries P, Gosselin F, Goebel R, Schyns PG (2009) Inverse mapping the neuronal substrates of face categorizations. Cereb Cortex 19:2428–2438. CrossRef Medline
Tanaka JW, Farah MJ (1993) Parts and wholes in face recognition. Q J Exp Psychol A 46:225–245. CrossRef Medline
Taylor MJ, Edmonds GE, McCarthy G, Allison T (2001) Eyes first! Eye processing develops before face processing in children. Neuroreport 12:1671–1676. CrossRef Medline
Tong F, Nakayama K, Moscovitch M, Weinrib O, Kanwisher N (2000) Response properties of the human fusiform face area. Cogn Neuropsychol 17:257–280. CrossRef Medline
Tsao DY, Livingstone MS (2008) Mechanisms of face perception. Annu Rev Neurosci 31:411–437. 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
Ullman S, Vidal-Naquet M, Sali E (2002) Visual features of intermediate complexity and their use in classification. Nat Neurosci 5:682–687. Medline
Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RB, Orban GA (2001) Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. Neuron 32:565–577. CrossRef Medline
Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. IEEE Comput Soc Conf Comput Vis Pattern Recognition 1:I511–I518. CrossRef
Walker-Smith GJ, Gale AG, Findlay JM (1977) Eye movement strategies involved in face perception. Perception 6:313–326. CrossRef Medline