Stimulus-dependent position sensitivity in human ventral temporal cortex

Rory Sayres[1], Kevin S. Weiner[1], Brian Wandell[1,2], and Kalanit Grill-Spector[1,2]
[1] Psychology Department, Stanford University, Stanford, CA 94305
[2] Neuroscience Program, Stanford University, Stanford, CA 94305

Abstract

Human ventral temporal cortex (VTC) contains high-level visual regions involved in face and object recognition. Strong position effects are reported in VTC, but the nature of these effects is unclear. We measured position sensitivity across visual cortex, including VTC and face-selective regions, using fMRI and a population receptive field (pRF) model. While fixating, subjects viewed slowly moving bar apertures containing either: phase-scrambled faces; a field of small faces; faces scaled with eccentricity; or parts of a single large face. While V1 responses to these stimuli are highly consistent, VTC demonstrates a striking stimulus-dependence. Estimated pRFs are systematically larger and centered further from the fovea for larger, less crowded face stimuli. Nevertheless, the relative position preference across voxels is maintained. Three distinct face-selective regions (IOG, pFus, and mFus) show increasing stimulus-dependence and foveal bias. These results suggest a processing hierarchy in VTC, combining position and stimulus information.

Introduction

Face recognition depends on a hierarchy of visual processing regions extending ventrally from primary visual cortex (V1) through ventral temporal cortex (VTC). Neural response properties change from early to later stages of the hierarchy in two key respects. First, there is decreasing retinal position sensitivity. Sizes of spatial receptive fields (RFs) increase, as measured using physiology (Boussaoud et al., 1991; Gross et al., 1969; Zoccolan et al., 2007) and functional magnetic resonance imagining (fMRI: (Amano et al., 2009; Dumoulin and Wandell, 2008)). Second, there is increased stimulus sensitivity. While neurons in early visual regions are selective for relatively simple features such as oriented edges (Hubel, 1995), single-unit responses in IT become highly selective for complex shapes and objects (Ito et al., 1995; Tovee et al., 1994; Zoccolan et al., 2007). Further, neuroimaging methods consistently identify face-selective regions, which exhibit a strong preferential response to faces over non-face stimuli (Kanwisher et al., 1997; Puce et al., 1995; Tsao et al., 2006). In humans, multiple, anatomically distinct face-selective regions have been identified along the inferior occipital gyrus (IOG) and posterior and mid-fusiform gyri (Grill-Spector et al., 2004; Pinsk et al., 2009; Tsao et al., 2008; Weiner and Grill-Spector, 2009)

A commonly-held view of the ventral visual processing stream posits that position information is discarded at the highest levels of this hierarchy (Goodale and Humphrey, 1998; Lueschow et al., 1994; Milner and Goodale, 1993; Ungerleider and Haxby, 1994). Under this model, objects and faces are represented in a largely position-insensitive manner (sometimes referred to as "position invariance"). Such position-insensitive representations are thought to underlie the robustness in our ability to recognize objects across retinal positions (Ellis et al., 1989; Thorpe et al., 2001). Several lines of evidence have been cited to support this view. Measurements in monkey IT cortex reported large RFs, of up to 50° or more (Boussaoud et al., 1991; Gross et al., 1969; Gross et al., 1972; Tovee et al., 1994). In humans, there is a greater level of fMRI-adaptation in VTC compared to earlier visual regions for objects presented at different positions (Grill-Spector et al., 1999). This has been interpreted as reflecting position–insensitive adaptation (Konen and Kastner, 2008). Similarly, changes in the distributed fMRI response to objects at different positions vary across regions. Although these changes are substantial, they are relatively more modest for anterior regions such as VTC, which has again been interpreted as reflecting position invariance (Schwarzlose et al., 2008). For these reasons, the view that neural representations in VTC are largely insensitive to retinal position predominates to this day.

However, a spate of recent evidence challenges the notion of position-insensitive representations in VTC. Certain VTC regions prefer particular eccentricities—face-

selective regions overlap with foveal representations and place-selective regions overlap with peripheral representations (Hasson et al., 2003; Levy et al., 2001). Responses in VTC are also strongly lateralized: neural responses in the right hemisphere are stronger for stimuli in the left versus right visual field, and vice versa (Grill-Spector et al., 1998; Gross et al., 1969; Niemeier et al., 2005; Sayres and Grill-Spector, 2008). This is a striking difference to our recognition abilities, which are comparable for objects an equal distance to the left or right of fixation (Thorpe et al., 2001). Further, position affects distributed patterns of response across cortex. The distributed VTC response changes markedly for objects at different visual field positions (Carlson et al., 2009; Schwarzlose et al., 2008). Similarly, more recent electrophysiology studies in monkeys reported that some IT neurons show strong sensitivity to stimulus position. These studies found large changes in neuronal firing rates for small changes in stimulus position, corresponding to RF sizes as small as a few degrees (DiCarlo and Maunsell, 2003; Op De Beeck and Vogels, 2000). Finally, behavioral studies of face adaptation effects show retinotopically specific effects (Afraz and Cavanagh, 2008), suggesting that face-processing mechanisms are position-sensitive. Together these results suggest that in both monkeys and humans there is considerable position sensitivity in VTC.

Our understanding of the magnitude and nature of VTC position effects is complicated by a large variability in the literature. A comprehensive set of measurements in monkey IT cortex showed a wide range of RF sizes, from 2.6 – 26° (Op De Beeck and Vogels, 2000). The small RFs observed in (DiCarlo and Maunsell, 2003; Op De Beeck and Vogels, 2000) differ from other measurements which reported large RFs in IT cortex (Boussaoud et al., 1991; Gross et al., 1969; Gross et al., 1972; Ito et al., 1995; Tovee et al., 1994). These differences may result from many factors, including the use of anesthetized versus awake animals, the fixation tasks and attentional controls used (Richmond et al., 1983) and measurements from functionally distinct regions within IT cortex. Finally, the stimulus used to determine position effects may be critical (DiCarlo and Maunsell, 2003; Gross et al., 1969). Studies reporting larger RFs tended to use large stimuli (>11°) to evoke neural responses (Gross et al., 1969; Gross et al., 1972; Tovee et al., 1994), while studies showing small RFs used smaller stimuli (0.6-3°) (DiCarlo and Maunsell, 2003; Op De Beeck and Vogels, 2000). This raises the possibility that position effects in VTC are stimulus-dependent. Despite its substantial implications for understanding ventral visual cortex, this hypothesis has not been directly tested, either in monkeys or humans.

Therefore, we measured retinotopic sensitivity in human face-selective cortex, and surrounding parts of ventral temporal cortex. We asked: (1) Can we quantitatively measure retinotopic sensitivity to face stimuli in VTC using fMRI? (2) Do position effects depend on the stimulus? (3) Does position sensitivity vary across face-selective regions?

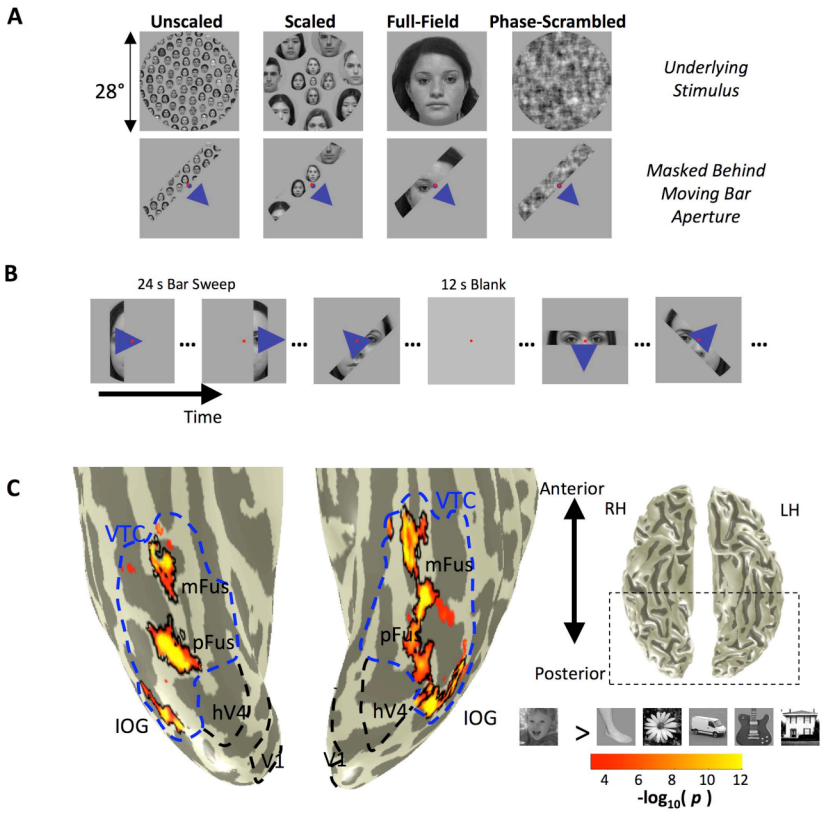


Figure 1. Experimental design. **(A)** Illustration of the four stimulus conditions in the main experiment. During the main experiment, the stimulus was a moving bar aperture that revealed one of four different types of underlying stimuli. Top row, examples of the underlying stimuli (L-R): phase-scrambled faces; small, crowded fields of faces which are unscaled with eccentricity; fields of faces which are scaled (2x) with eccentricity; and full-field views of a single face. Bottom row, example image frame from a single time point showing the appearance of each stimulus when masked by a bar aperture. The red dot indicates the fixation point (size is exaggerated for visibility; actual size ~0.5°); the arrow indicates the direction of motion of the bar aperture. **(B)** Illustration of the movement of the bar aperture during one run. The bar moved smoothly across visual space in eight different directions (4 cardinal directions plus 4 diagonals), completing one sweep every 24 seconds. 4 times during the scan, the bar would disappear, leaving a mean-luminance screen for 12 seconds. Subjects fixated on the central red point) and performed a color discrimination task during each run. **(C)** Face-selective regions in ventral-temporal cortex,. *Left:* inflated view of the ventral surface of subject S1, with a statistical contrast map of (faces) > (objects) overlaid at $p < 0.001$, uncorrected. V1 and V4 are outlined in cyan. The three face-selective regions are outlined in black; from posterior to anterior: inferior occipital gyrus (IOG), posterior fusiform (pFus), and middle fusiform (mFus). The blue outline indicates an anatomical definition of ventral temporal cortex (VTC); Right: ventral view of the cortical gray matter surface, with the imaged region of ventral temporal cortex indicated by the dotted rectangle.

We focus on three face-selective regions within VTC (Pinsk et al., 2009; Weiner and Grill-Spector, 2009), in order to examine whether there are regional differences in position effects. These regions were identified via independent localizer scans. We refer to these regions by their anatomical locations: they are in and around the inferior occipital gyrus (IOG), posterior fusiform gyrus (pFus), and mid-fusiform gyrus (mFus). Finally, we also analyzed responses across the anatomically-defined ventral temporal cortex (VTC), excluding face-selective regions, in order to measure whether our effects were specific to face-selective cortex.

Results

We utilized a new methodology developed by Dumoulin and colleagues that estimates the population receptive field (pRF) from fMRI data (Dumoulin and Wandell, 2008). This method provides a representation in visual space (modeled by a 2-D Gaussian) that most effectively activates each fMRI voxel. We employed novel mapping stimuli containing different types of faces (**Figure 1**). In each of our four experiments a moving bar aperture slowly swept across the visual field in different directions, turning off periodically to allow a measure of the baseline response. On different experimental runs, the bar revealed either **(1)** a field of small (~2°) faces spaced closely together ("Unscaled Faces"); **(2)** a field of faces spaced further apart and scaled with eccentricity ("Scaled Faces"); **(3)** a single large face spanning the extent of the stimulus ("Full-field Faces"); or **(4)** a phase-scrambled version of the scaled faces, with the same spatial frequency content but without the appearance of a face ("Phase-scrambled Faces"). The motion of the bar and the extent of visual space stimulated at each time point was constant across each of the four stimulus types.

Face-containing stimuli activate ventral temporal cortex in a position specific manner

Presenting intact faces behind a moving bar aperture produces strong, retinotopically-sensitive BOLD responses in some regions of ventral temporal cortex. **Figure 2** displays the time courses from a typical face-selective voxel in the middle fusiform gyrus (mFus) face-selective region (**Figure 1C** shows the ROI**)** from an example subject. When the subject viewed the scaled faces mapping stimulus behind the moving bar, this voxel showed a robust modulation as the retinal position of the bar aperture varied (**Figure 2A-dashed**). We model the BOLD response by fitting a population receptive field (pRF) to the data. This pRF is a circular 2D Gaussian function in visual space, which describes the extent of visual space for which visual stimuli effectively elicit a response. When the bar aperture crosses the voxel's pRF the signal increases and when the aperture moves away and does not overlap the pRF, the signal decreases. In this example, the pRF

which best fits this response is centered 3.8° from the fovea in the lower left visual field and its half-width half-max is 6.5°. The predicted response (**Figure 2A-solid**) fits the measured response (**Figure 2A-dashed**) well, explaining 67% of the response variability. The same pRF estimate also provides a good fit for the voxel's response during an independent set of runs with the same stimulation (**Figure 2B**), explaining 71% of the response variance.

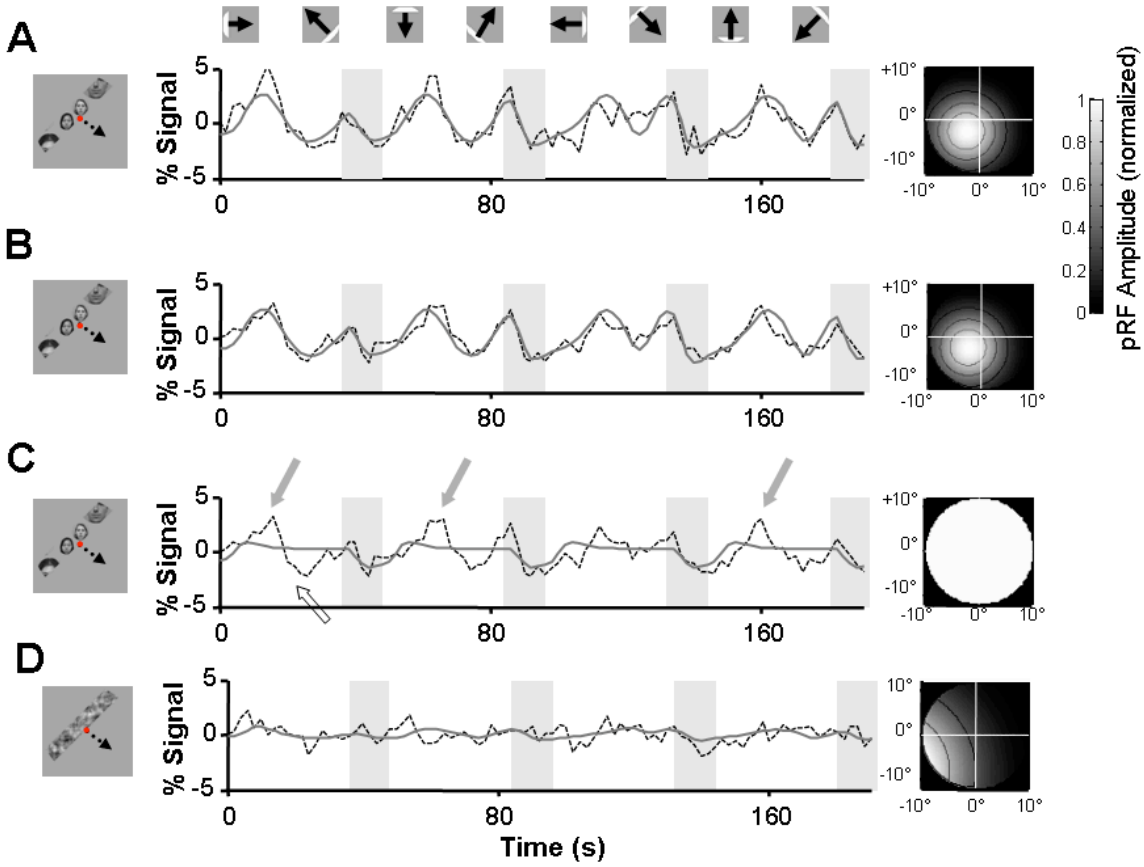


Figure 2. Face stimuli excite ventral-temporal cortex in a retinally specific manner. The illustrated time courses were taken from a representative voxel in the mFus ROI from subject S2. **(A)** The response to a face stimulus modulates as the bar aperture moves across the visual field. This demonstrates that the excitation is retinally specific. **(B)** The response to the same stimulus during an independent set of experimental runs. The same pRF solved from the data in (A) is used to model this separate response, and produces a good fit (71% variance explained), demonstrating reproducibility of the retinally-specific response. **(C)** The data are not well-fit using a large, uniform pRF that shows no significant retinotopic sensitivity. This model predicts response modulations only during the four blank periods, when the bar aperture disappeared (*shaded regions*). Although those parts of the response are well-fit, many other response peaks and troughs are not predicted by an invariant model (*arrows;* 22% variance explained). **(D)** The response to a phase-scrambled face is noisy and the best fit pRF explains 16% variance of this time course. For each row,

the measured BOLD time courses is shown in dashed black and the predicted time courses from the estimated population receptive field (pRF) is shown in gray. Zero reflects the mean of the time courses. The pRF is shown in visual space to the right, with circles denoting the contours when the pRF is at 75%, 50%, and 25% of the peak amplitude. The pRFs are computed in units of predicted % BOLD response per square degree, although for display purposes the pRFs are shown with the same peak height. The direction of the bar sweeps over the course of the run is indicated at the top of the figure.

The strong modulations reflect substantial position sensitivity. One way to assess the position sensitivity is to compare the responses with the expected pattern of a pRF with complete position-invariance. In that case, the expected response is a constant, only returning to baseline during the four mean-luminance periods when the face stimuli are removed. The idealized "position-invariant" prediction fails to predict the voxel's response modulations during the bar sweeps (**Figure 2C**, arrows). Overall, the position-invariant model predicts far less of the response variance in the VTC voxels than the pRF model (**Supplemental Figure 1**; left column of histograms).

Position dependent responses are more clearly observed with face-containing mapping stimuli than with phase-scrambled faces. For the same fusiform voxel, the BOLD response to a phase-scrambled face is low and noisy (**Figure 2D**). The best-fit pRF model only explains a small portion (16%) of the response variance. We measure the level of position sensitivity by measuring the variance explained by the pRF model for each stimulus. Across VTC, the variance explained by the pRF model is significantly higher for face-containing stimuli than phase-scrambled face stimuli (**Figure 3**), and there is a greater degree of position sensitivity for face-containing bars than phase-scrambled face bars. These effects are shown on a smoothed cortical surface of VTC: whereas phase-scrambled faces elicit weak responses that are not well-fit by the pRF model (**Figure 3A-left**), for Scaled Faces a broad extent of VTC shows a high degree of variance explained (**Figure 3A-right**), both within and around face-selective regions (*black outlines*).VTC exhibits substantial increases in variance explained for each of the three face-containing mapping stimuli relative to phase-scrambled (**Figure 3B**). These increases are statistically significant ($p < 0.01$, one-tailed T test, Bonferroni corrected) for pFus and mFus, as well as more broadly in ventral temporal cortex. (In IOG we found a similar, albeit non-significant trend.) By contrast, the variance explained by the pRF model is similar for all stimulus types in V1.

In addition to the variability in the time series fit by the model, we also examine whether the response amplitude increases for intact compared to phase-scrambled faces. We compare only the subset of voxels in each ROI with at least 15% variance explained for phase-scrambled faces, to restrict us to the situation in which we are reliably measuring signals for both intact and scrambled faces. Among these voxels, the amplitude of

response to each of the intact face stimuli is consistently higher than for phase-scrambled faces (all $p < .0008$, paired one-tailed T test), in pFus, mFus, and across the anatomical extent of the VTC. For IOG, both Scaled and Full-Field faces produced significantly higher amplitudes than phase-scrambled ($p < 10^{-14}$), but not Unscaled Faces ($p = 0.96$).

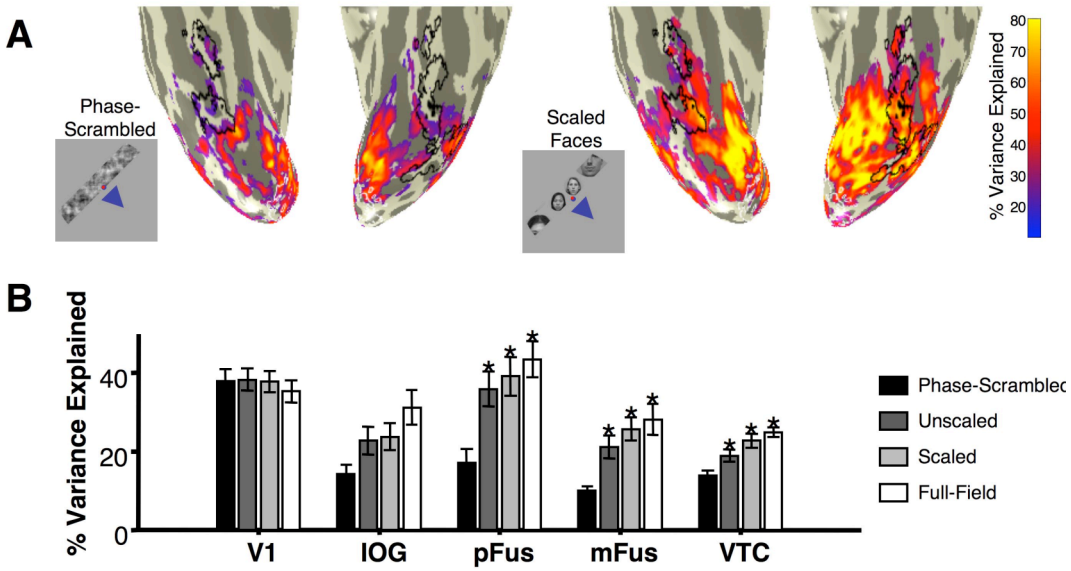


Figure 3. Retinally-specific modulation, as measured by the proportion variance explained, increases across ventral temporal cortex for intact versus phase-scrambled faces. (A) The model fit (proportion variance explained in each voxel's time courses) is displayed on inflated representations of the ventral cortical surface of subject S1. The variance explained is thresholded at 20%. The model fit is fairly poor across ventral temporal cortex, when measured using phase-scrambled face stimuli (*left*), especially within face-selective regions (*black outlines*). In the same subject, a much greater amount of variance is explained in the response to the Scaled-faces mapping stimulus (*right*), both within and nearby face-selective regions. (B) The mean variance explained measure across each of the four mapping stimuli for V1, the three face-selective regions (IOG, pFus, and mFus), and the anatomically-defined extent of VTC as a whole. Asterisks indicate significantly ($p < 0.01$, Bonferroni corrected) higher variance explained as compared to the phase-scrambled stimulus.

Retinotopic sensitivity in ventral temporal cortex depends on the mapping stimulus

Given that face-containing stimuli evoke a position sensitive response in VTC, we now ask whether position sensitivity varies with the type of face stimulus. **Figure 4** shows the time courses of response to three different face-containing stimuli from two example voxels in the same subject. Typical V1 voxel (panel A) and a typical voxel from the face-selective mFus region (panel B). The responses in the V1 voxel are highly consistent

across the three stimuli (**Figure 4A**; pairwise correlations R range from 0.88 to 0.90). Because the responses are the same, the best-fit pRF estimate is essentially unchanged despite changing the stimuli. The pRFs are consistently located near the left horizontal meridian (eccentricity=6.6, 6.2, and 6.6°, respectively, for the three stimuli) and are consistently small (HWHM=2.4, 2.2, and 2.1°). The small pRF size explains the sharp temporal profile of responses: as the bar sweeps past the small pRF, the BOLD response first increases rapidly, and then returns to baseline.

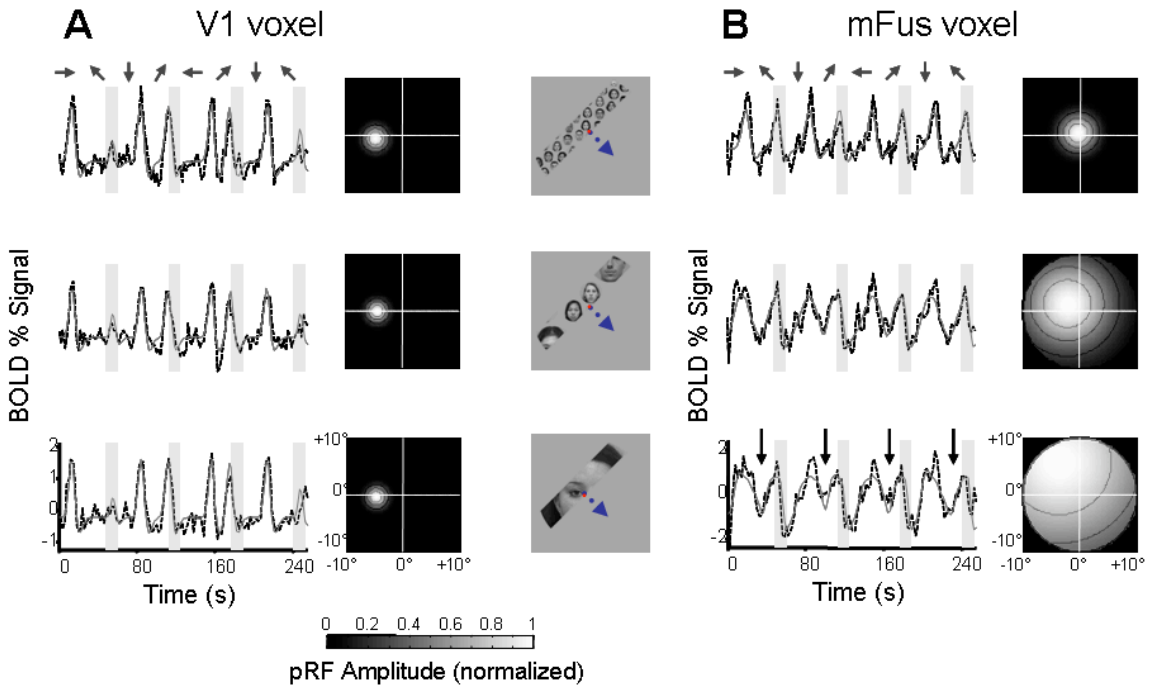


Figure 4. Retinotopic sensitivity in mFus, but not V1, depends on the stimulus.

(A) Time series and pRF fits for a typical V1 voxel (subject S2). From top to bottom the face stimuli were: Unscaled, Scaled, and Full-field. The time courses, and hence the estimated pRFs, are constant across the face stimuli. The estimated pRF eccentricities (size) are 6.6 (2.4)°, 6.2 (2.2)° and 6.6 (2.1)°. Zero reflects the mean of the time courses. **(B)** Time series and pRF fits for a typical mFus voxel to the same three types of face stimuli. The duration of the positive BOLD signal in the time courses increases as face size increases. Consequently the pRF size and eccentricity change, becoming larger and more eccentric. The estimated pRF eccentricities (size) are 0.8 (3.7)°, 3.4 (9.1)°, and 9.4 (19.9)°. The response to Full-Field faces is very broad, but exhibits a decrease when the bar is in the ipsilateral visual field (*arrows*). The variance explained by the pRF model for both of these voxels exceeds 64% in all cases. Time series plotting conventions and other details as in **Figure 2**.

The V1 responses stand in stark contrast to the VTC responses. An example mFus voxel (**Figure 4B**) responds with a narrow response to the Unscaled Faces stimulus, resulting in a pRF that is fairly small (3.7°; top row) and close to the fovea

(eccentricity=0.8°). The response to the Scaled faces is substantially different (middle row), becoming more prolonged. Consequently, the best-fit pRF is substantially larger (8.2°) and displaced from the fovea (eccentricity = 3.0°). Finally, when Full-field faces that fill the display are presented behind the same moving bar aperture, the BOLD response becomes even broader (bottom row) and reaches the minimum only during the blank periods indicated by the gray zones. The resulting pRF estimate is very large (19.9°) and displaced even farther from the fovea (eccentricity=9.4°).

Although the responses to Full-field faces are very broad, and result in a very large pRF, this voxel does not respond to stimuli across the entire visual field. The pRF predicts a substantial contralateral bias: stimuli displaced equally to the left or right of the fovea should produce BOLD response levels. This visual field asymmetry is visible in the time series: as the bar passes into the ipsilateral visual field, the response decreases (shown by the arrows). The idealized "position-invariant" prediction illustrated in **Figure 2C** does not explain the response to Full-field faces well (**Supplemental Figure 1**).

The effects illustrated for this example voxel are typical across voxels and subjects, That is, changing the stimulus seen through the bar aperture has a significant impact on the VTC responses. In V1, by contrast, there is no systematic change in the response. For example, within V1 across all subjects, the time courses of response to the three different stimuli are highly correlated, with a mean R of 0.66 (standard deviation 0.11 across all voxels). By contrast, in VTC the mean correlation in responses across voxels is significantly lower (0.52 ±0.09, $p < 10^{-15}$, paired-sample T test). The V1 correlations are also significantly greater than in each of the three face-selective regions within VTC with the lowest correlation in the mFus (IOG: mean R = 0.59 ± .12, p = 0.001; pFus: mean R = 0.52 ± .12, $p < 10^{-13}$; mFus: mean R = 0.41 ± .14, $p < 10^{-12}$). We note that the in VTC the correlations across-stimuli, while lower than V1, are still significant and positive. We will discuss this further later.

Regional specificity of stimulus-dependence

We now broaden the analysis from example voxels to the cortical surface. **Figure 5** shows maps of the eccentricity and size of the estimated pRFs in an example subject's right ventral posterior cortex. We focus our analyses on these maps because the polar angle estimates in VTC were less consistent within and across subjects than the size or eccentricity estimates (**Table 2**).

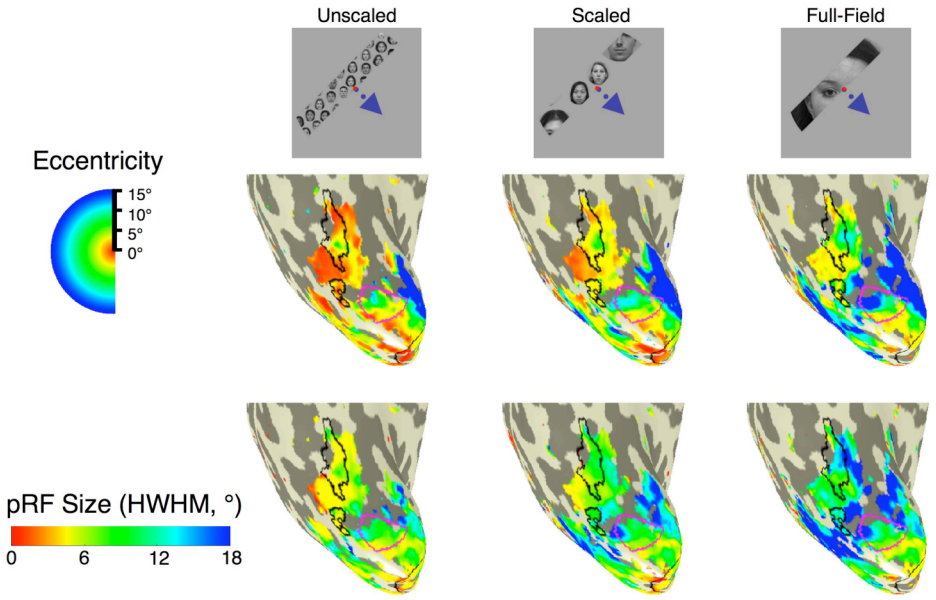


Figure 5. Stimulus-dependent eccentricity and pRF size maps in ventral temporal cortex. An inflated representation of the right cortical hemisphere is shown for subject 2, overlaid with maps of estimated pRF eccentricity (*top*) and size (*bottom*). Voxels are colored if the pRF model explained at least 15% of the variance of their time courses. Solid black outline indicate the pFus and mFus face-selective regions in this subject determined from an impendent localizer. Dotted black outline indicates hV4.

Eccentricity maps in early visual regions derived from the face-mapping stimuli are consistent with the known retinotopic organization in these regions. For example, moving from posterior to anterior along hV4 (**Figure 5-dotted outline**), eccentricity representations progress smoothly from foveal, through parafoveal to peripheral eccentricities (top row). This progression is mirrored by a corresponding increase in pRF size (bottom row), which is known to correlate with eccentricity across visual field maps (Amano et al., 2009; Winawer et al., 2010). The same pattern is visible along the medial surface in early visual regions V1-V3 (**Supplemental Figure 2**). These patterns are consistent across the three face-containing mapping stimuli.

Along the lateral aspect of the ventral surface, near face-selective regions (**Figure 5-solid black outline**), the eccentricity and size maps vary with the stimulus type. The eccentricity map for Unscaled Faces (**Figure 5-left**) is consistent with the coarse eccentricity topography reported in (Hasson et al., 2002; Levy et al., 2001; Malach et al., 2002): voxels in face-selective regions respond maximally to foveal eccentricities (red) while the more medial regions which contain place-selective regions prefer peripheral eccentricities (blue). However, the eccentricity of face-selective voxels changes when we use Scaled and Full-field Faces. When we compare the Unscaled map to the map for

Scaled faces, we find that the foveal bias (shown by the concentration of red) is substantially reduced (a greater amount of yellow parafoveal eccentricity).This effect is even more pronounced for the Full-Field faces: the eccentricity preference of these face-selective voxels is as far as 10° or more from the fovea. These maps illustrate a range of eccentricities within face-selective regions. Similar stimulus-dependent effects occur for the size maps. PRF sizes in face-selective regions are smallest when estimated with Unscaled Faces and progressively increase with Scaled and Full-Field faces, respectively.

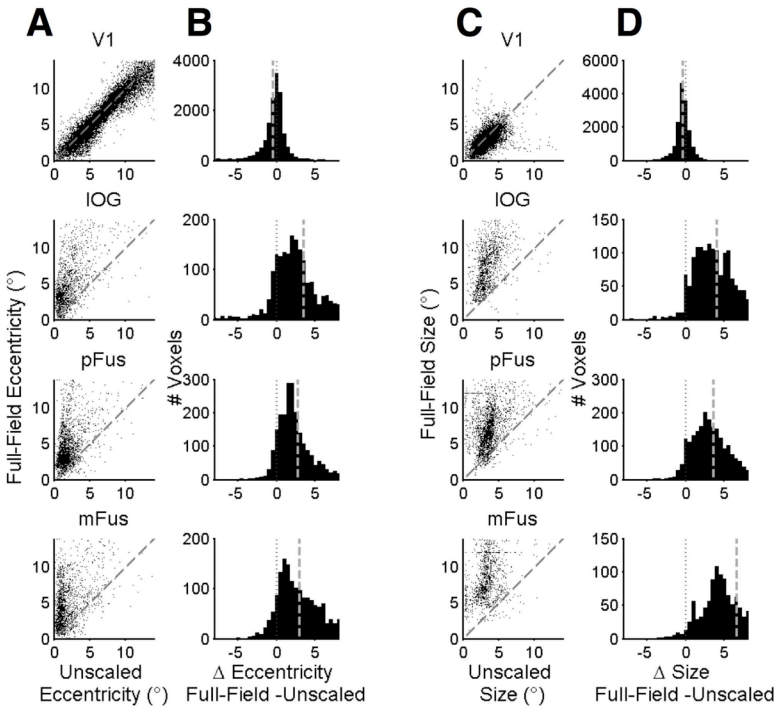


Figure 6. pRF estimates of size and eccentricity are larger for Full-Field vs. Unscaled Faces. (A) Scatter plots showing the joint distribution of pRF eccentricity estimates from Full-Field versus Unscaled Faces, for V1 (first row), IOG (second row), pFus (third row), and mFus (fourth row). Each data point represents a single voxel; data are taken from all subjects (N=7). Whereas V1 voxels are clustered around the X=Y line, they are significantly above this line for the three face-selective ROIs. This demonstrates a stimulus dependence for eccentricity estimates in the face regions, but not V1. **(B)** The distribution of the difference in the eccentricity estimates between the Full-Field and Unscaled Faces stimuli for each ROI (across all voxels and subjects). The dashed line shows no difference and the solid gray line indicates the mean (these values are also provided in Table 1). **(C)** Scatter plots of pRF size estimates from Full-Field and Unscaled Faces, same conventions as **(A)**. **(D)** Distributions of the differences in size estimates between the two mapping stimuli, same conventions as **(B)**.

The size and eccentricity maps illustrate that the dependence of pRF estimates on the mapping stimulus is a regionally-specific effect. We observe stimulus-dependence in ventral temporal cortex, but not in early visual cortex. This regional contrast can be quantitatively measured across subjects by comparing the distributions of pRF size and eccentricity for pairs of different stimuli (**Figure 6**).

In V1, size and eccentricity estimates are consistent across stimuli. The scatter plot in the top row of **Figure 6a** plots the eccentricity estimates from the Full-Field Faces stimulus against those from Unscaled Faces across all subjects. As can be seen, the data points cluster about the X=Y line, indicating that the pRF estimates derived from these two stimuli are consistent with one another across voxels**.** There is no significant difference in eccentricity estimates across mapping stimuli, since the mean difference in the eccentricity estimate is not significantly different than zero (**Figure 6b, top**). Likewise, the size estimates in V1 are very similar for the two face stimuli (**Figure 6c,d**), and the mean difference in size estimates between the two stimuli is centered around zero.

The fMRI time courses in much of ventral temporal cortex vary with changes in the size and density of the face stimuli. Consequently, the estimated pRF sizes and eccentricities estimated from Full-Field faces are systematically larger than from Unscaled Faces. Within each face-selective ROI, the distribution of eccentricity estimates is larger for Full-Field than Unscaled Faces. Consequently, voxels lie above the X=Y line in rows 2-4 of **Figure 6a**. The distribution of eccentricity differences across voxels and subjects for each face-selective ROI is significantly positive ($p < 10^{-30}$ for each distribution, two-tailed T-test, **Table 1**). Likewise, we find similar increases in the size estimates for Scaled - Unscaled and Full-Field - Scaled comparisons (**Figure 6** and **Table 1)**. Notably, the greatest increase is between the Full-Field - Unscaled Faces; the increase in the estimated pRF size is larger than the increase in the estimated eccentricity.

ROI	Eccentricity Shift (°)	Size Shift (°)
Scaled - Unscaled Faces		
V1	-0.03 (2.61)	-0.32 (1.25)
hV4	**0.45 (3.25)**	**0.53 (2.12)**
IOG	**1.02 (4.64)**	**1.94 (3.54)**
pFus	**1.65 (2.58)**	**1.91 (2.25)**
mFus	**1.81 (3.05)**	**3.37 (3.28)**
VTC	**1.24 (4.26)**	**1.96 (2.94)**
Full-Field - Unscaled Faces		
V1	-0.47 (2.64)	-0.38 (1.22)
hV4	**2.26 (4.01)**	**1.29 (2.41)**
IOG	**3.49 (5.18)**	**3.83 (3.30)**
pFus	**2.72 (3.01)**	**3.60 (3.01)**
mFus	**2.94 (3.28)**	**6.59 (4.29)**
VTC	**3.09 (4.79)**	**3.96 (3.94)**
Full-Field - Scaled Faces		
V1	-0.44 (2.76)	-0.07 (1.12)
hV4	**1.81 (3.82)**	**0.76 (2.25)**
IOG	**2.46 (4.48)**	**1.89 (3.44)**
pFus	**1.07 (3.08)**	**1.69 (2.69)**
mFus	**1.14 (4.10)**	**3.22 (4.20)**
VTC	**1.85 (4.31)**	**2.00 (3.40)**

Table 1. Magnitude of the differences in the pRF estimates for different pairwise comparisons of mapping stimuli. The difference is the mean value (and standard deviation) across all voxels, hemispheres, and subjects. The data are presented in three groupings: Scaled – Unscaled, Full-Field – Unscaled and Full-field – Scaled. Each grouping summarizes the difference between the estimated eccentricity (left) and size (right) for two different face-containing mapping stimuli. Amplitudes in bold are significantly different from zero (two-tailed paired T test across voxels, $p < 10^{-20}$ for all significant comparisons)

Relative position preference is preserved across mapping stimuli

The absolute pRF sizes and eccentricities change across stimuli, but the relative values are preserved. Some voxels respond preferentially to more foveal eccentricities than other nearby voxels, which prefer more peripheral stimulation, even though the absolute eccentricity preference of these voxels is sensitive to the size and crowding of face stimuli. This trend is visible in the scatter plots (**Figure 6**), where voxels in face-selective regions tend to cluster around a line with slope greater than 1.

We assessed the consistency in the relative position preferences across VTC by correlating the size and eccentricity maps illustrated in **Figure 5**. We computed these correlations for each pairwise comparison between two face-containing mapping stimuli (**Table 2**).

ROI	Eccentricity Correlation (R)	Size Correlation (R)
Scaled - Unscaled Faces		
V1	**0.91 (0.05)**	**0.63 (0.19)**
hV4	**0.80 (0.13)**	**0.66 (0.16)**
IOG	**0.56 (0.34)**	**0.37 (0.32)**
pFus	0.21 (0.37)	0.24 (0.33)
mFus	**0.34 (0.36)**	**0.37 (0.33)**
VTC	**0.72 (0.20)**	**0.59 (0.19)**
Full-Field - Unscaled Faces		
V1	**0.91 (0.05)**	**0.66 (0.16)**
hV4	**0.82 (0.12)**	**0.68 (0.15)**
IOG	**0.56 (0.24)**	**0.42 (0.19)**
pFus	**0.38 (0.28)**	**0.26 (0.30)**
mFus	0.14 (0.28)	**0.33 (0.27)**
VTC	**0.65 (0.19)**	**0.36 (0.17)**
Full-Field - Scaled Faces		
V1	**0.90 (0.05)**	**0.65 (0.15)**
hV4	**0.84 (0.09)**	**0.74 (0.10)**
IOG	**0.58 (0.27)**	**0.42 (0.33)**
pFus	0.33 (0.43)	0.33 (0.41)
mFus	0.21 (0.35)	0.30 (0.38)
VTC	**0.73 (0.16)**	**0.53 (0.15)**

Table 2. Correlations between pRF estimates from the different mapping stimuli. The data are presented in three grouping, representing different pairwise correlations between the three face-containing mapping stimuli. The correlations were computed across voxels within each hemisphere. The mean correlation coefficient across hemispheres and subjects (N=14 hemisphere) is shown, with the standard deviation in parentheses. Correlations marked in bold are significantly greater than zero across hemispheres (N=14, one-tailed T test against zero, $p < 0.01$, Bonferroni corrected).

The coarse eccentricity topography is well-preserved across the anatomical extent of the entire VTC (mean $R \geq 0.72$ across comparisons). That is, despite the change in the estimated eccentricity in the face-selective regions, the lateral aspect of the VTC exhibits a consistently more foveal preference compared to medial VTC (**Figure 5**). The pRF size maps also show positive, but significantly lower correlations across different face stimuli compared to the eccentricity correlations (mean $R \geq 0.37$ across comparisons for VTC; $p < 10^{-5}$, one-tailed paired T test). Within face-selective ROIs (IOG, pFus and mFus), the correlations are largely positive, but less consistently than across the whole VTC. Among the three face-selective ROIs, the relative eccentricity across voxels within the more posterior IOG region shows a greater consistency across mapping stimuli,

compared to the more anterior pFus and mFus fusiform ROIs and also compared to the estimated pRF size. This may reflect the larger eccentricity range in IOG, compared to the strong foveal bias in pFus and mFus (**Supplemental Figures 3,4**, especially for the Scaled faces.

These results relate to previous reports of a coarse eccentricity bias across VTC (Hasson et al., 2002; Levy et al., 2001; Malach et al., 2002). Consistent with these reports, we observe an enhanced representation of the fovea across face-selective voxels, especially in the pFus and mFus ROIs. However, different from previous reports, our data indicate that not all face-selective voxels respond maximally to foveal stimuli (**Supplemental Figures 3,4**). Instead, a significant number of voxels respond maximally to faces at more peripheral positions. In the IOG, in particular, we find many voxels that prefer peripheral stimulation (>10°) even for Unscaled face stimuli.

A hierarchical sequence of position and stimulus effects across ventral temporal cortex

There appear to be substantial qualitative differences among the three face-selective regions IOG, pFus, and mFus in terms of their position sensitivity, foveal bias and coverage of the visual field within each of these regions (**Figure 6a** and **c** and **Supplemental Figure 4**). For example, We computed maps of the pRF coverage, which take into account both position of the pRF centers and their size across all voxels in an ROI (Amano et al., 2009; Winawer et al., 2010).These coverage maps reveal strong regional differences in visual field coverage (Figure 7). Established visual field maps, such as the hemifield map in V1 and quarter field in V3v, show consistent coverage of the expected extent of visual space for all three face stimuli. By contrast, each of the three face-selective ROIs exhibit different levels of visual field coverage depending on the mapping stimulus.

Among the different face-selective regions, we also see differences in visual field coverage. For all mapping stimuli, the IOG covers a substantial portion of the contralateral hemifield. In contrast, mFus shows a consistent foveal representation, and its coverage extends substantially into the ipsilateral visual field across all mapping stimuli. The degree of stimulus-dependence in visual field coverage also varies across face-selective ROIs. Changes in coverage across stimuli are fairly subtle for IOG, more substantial in pFus, and dramatic in mFus. There is also a greater coverage of the ipsilateral visual field for large stimuli, in part because of the substantial increases in pRF size (**Table 1**).

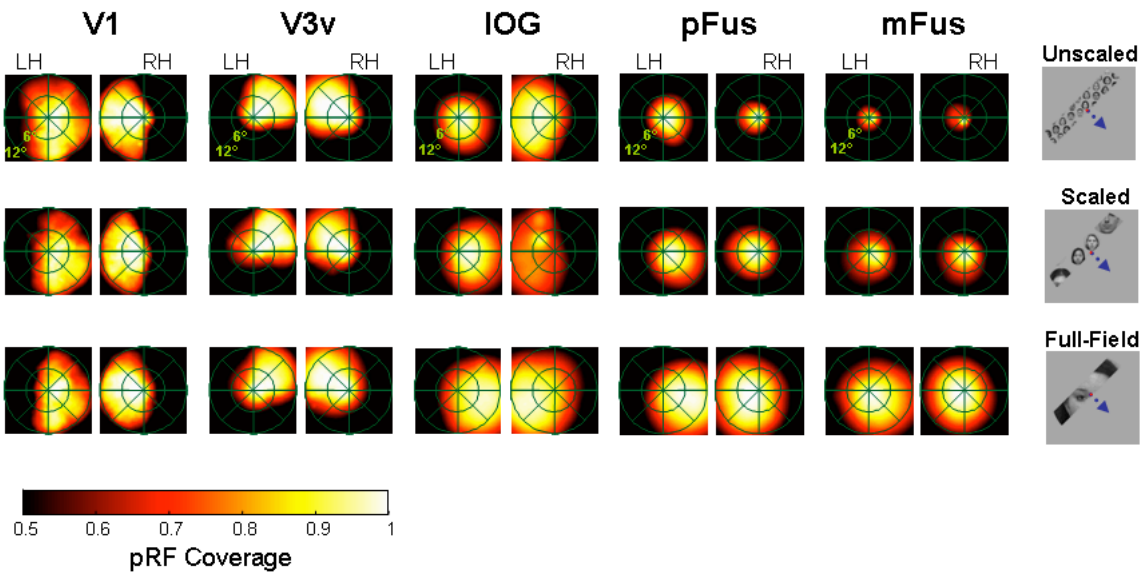


Figure 7. PRF coverage maps in visual space for V1 and face-selective regions. Each plot shows the mean coverage of the visual field, determined by the position and size of pRFs across voxels from all subjects (N=7). The coverage is estimated using the maximum projection method (for details, see (Amano et al., 2009)). *Left to Right*, coverage in V1, ventral V3, inferior occipital gyrus, posterior fusiform, and middle fusiform regions. *Top row*, coverage for Unscaled Faces; *Middle,* Scaled faces; *Bottom*; Full-field face. For each ROI, the left hemisphere coverage maps are shown in the left column, and the right hemisphere coverage maps are in the right column. Whereas V1 and V3v exhibit a consistent coverage of the contralateral visual field, IOG, pFus and mFus exhibit a differential level of coverage depending on the face stimulus.

These coverage maps predict differences in position sensitivity across the three face-selective ROIs. Specifically, we would expect an increasing foveal bias as we progress from the IOG, to pFus, to mFus, as well as an increasing response to faces presented ipsilaterally to each hemisphere. We tested these predictions by examining the responses to face images from a separate experiment (in which the same subjects viewed 2.5° faces that were presented in blocks, at the fovea, 4° to the left and 4° to the right of fixation (see (Sayres and Grill-Spector, 2008) for details).

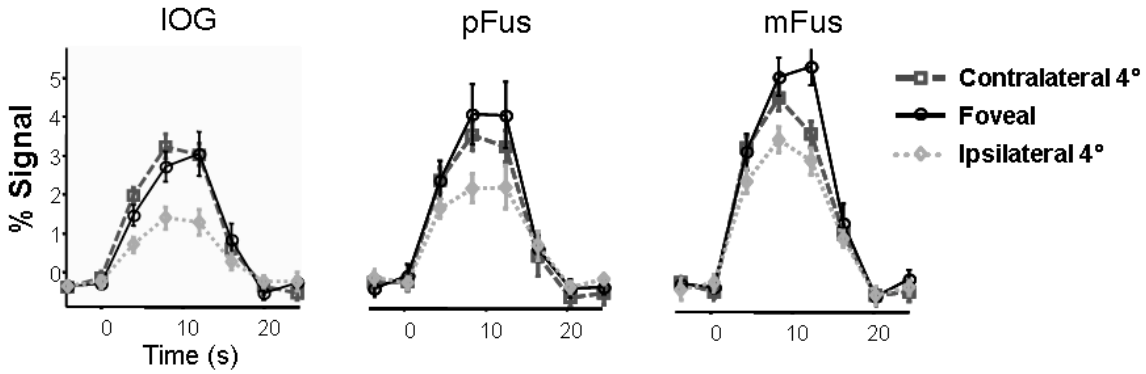


Figure 8. A hierarchical sequence of position sensitivity across face-selective ROIs. Face images subtending 2.5° were presented one face at a time in 12-s blocks, at three positions: foveally, 4° to the left, and 4° to the right of fixation. Data were combined across hemispheres, sorting stimuli by whether they were foveal (*black*), contralateral (*dark gray*) or ipsilateral (*light gray*) to a given hemisphere. The mean peri-stimulus time courses are shown for each of the face-selective ROIs: inferior occipital gyrus (*left*), posterior fusiform (*middle*), and mid-fusiform (*right*). Error bars represent standard error of the mean across hemispheres (N=10).

Consistent with these coverage maps, the mean response to faces at different positions exhibits a increasing foveal bias, and increasing ipsilateral response, from IOG to pFus to mFus. The pattern of response in our IOG region is different from that reported in (Hasson et al., 2003). We do not find evidence for a foveal bias in this region; instead there is a significant contralateral bias and no difference between responses to foveal and contralaterally presented faces ($p = 0.41$, paired two-tailed T test on beta coefficients across hemispheres). For mFus, foveal responses are significantly higher than contralateral responses ($p = 0$ .02). It should be noted that the mean responses across these ROIs reflect the summation over a range of voxels within each ROI, which have heterogeneous position sensitivity. The comparable response to foveal and contralateral faces in the IOG reflects the contribution of separate subsets of voxels, some which prefer more foveal and others that prefer more parafoveal retinal positions.

Overall, there are substantial differences in coverage of the visual field across face-selective regions. These differences suggest a hierarchy among the three regions, in which sensitivity across the visual field becomes stronger for foveal stimuli, less lateralized to the contralateral visual field, and more sensitive to the nature of the face stimulus.

Discussion

We present for the first time quantitative measurements of population receptive field properties across ventral temporal cortex, including within face-selective regions. These results demonstrate several important findings. First, face-containing stimuli can produce robust, retinotopically-specific fMRI responses across ventral temporal cortex. Second, there is a profound dependence on the type of face stimulus used to elicit responses; this stimulus dependence is not observed in V1. While there a large stimulus-dependent changes in absolute position sensitivity, the relative position preference across the cortical surface is preserved across mapping stimuli. Third, the degree of stimulus dependence, as well as the coverage of visual space, differs between the three distinct face-selective regions, IOG, pFus, and mFus. These combined position and face-selective responses have implications for our understanding of the functional organization of ventral temporal cortex, and practical implications for measuring responses in these regions.

Position Effects in VTC Depend on the Stimulus

While VTC has significant sensitivity to stimulus position, this position sensitivity (and consequently the estimated pRF) depends on the stimulus. Face-selective voxels require an effective stimulus: they exhibit weak or no position sensitivity for phase-scrambled faces with the same contrast and spatial frequency distribution as faces, but without the appearance of a face (**Figures 2-3**). Position sensitivity in VTC voxels also depends on the type of face stimulus within the bar, not just the bar's position. While VTC voxels respond in a reproducible and retinally-specific manner to each of the three face-containing stimuli, the time course of response varies markedly across the three stimulus types (**Figure 4**). The effects of changing the stimulus content within the bar were pronounced in VTC, but did not affect V1 responses or its estimated pRFs. This stimulus-dependent position sensitivity reflects a basic difference in response properties between early and high-level visual cortex.

There are several possible interpretations of these stimulus effects. One possibility is that the different stimuli drive responses in distinct neural subpopulations within VTC. Different neurons may be tuned to faces of different sizes, while at the same time having measurably distinct receptive fields. Similar joint-tuning effects have been reported in monkey IT cortex (Ito et al., 1995). Joint tuning to size and position may be a useful property in developing neural representations of objects. For instance, it may facilitate recognition of multiple objects across the visual field, by allowing for non-overlapping representations of different-sized objects presented together (see (Li et al., 2009)).

If the different pRF estimates derive from separate neural subpopulations, this implies that the same stretch of cortex may contain multiple overlapping representations of visual space, each for a different neural subset (similar to that indicated in **Figure 7**). This may present complications for measuring functional organization, such as visual field maps, in these regions. It suggests that the different pRF estimates we derived are each accurately reflecting some aspect of neural responses in VTC. A full characterization of response properties, in this case, may require a multi-dimensional representation: for instance, measuring BOLD responsiveness as a function of visual position and face size, rather than a single two-dimensional position pRF.

An alternative interpretation is that the different stimuli drive the same neuronal population differentially. That is, the effects we observe with fMRI are due to differential responses at the single-neuron level. The data predict that neural RF sizes will be smaller and centered closer to the fovea when estimated with Unscaled versus Scaled faces (or Full-field faces). Some data from single neuron recordings in awake behaving macaque monkeys support this interpretation. Op De Beeck and Vogels (Op De Beeck and Vogels, 2000) illustrate that the effects we report here can occur in individual neurons in macaque IT cortex. The authors presented complex Fourier mapping stimuli at two sizes (3.3° and 6.6°) in different visual field locations while the monkey fixated. Across the 13 IT neurons measured, they reported a significant increase in the RF size and a shift in its center when measured with the larger stimulus. Future physiological recordings can elucidate these possibilities by performing measurements similar to those reported here, but at the single neuron level.

Whether the neural responses underlying these BOLD measurements are heterogeneous or homogenous in nature, fMRI measurements of position effects in VTC will be sensitive to the stimulus used. We next consider what aspects of our stimuli may be driving the observed stimulus-dependent position effects.

What stimulus factors affect position sensitivity in VTC?

To keep the retinal extent of visual stimulation constant across experiments, we co-varied several factors across stimuli, including the number, crowding, and size of faces. Since each of these can affect neural responses in primate ventral temporal cortex (Ito et al., 1995; Zoccolan et al., 2005), multiple factors may drive the observed interaction between stimulus content and position sensitivity. Our measurements do not separate the relative contribution of these effects, but it is likely that each plays a role in driving VTC responses. We consider below three factors that likely modulate VTC position effects: face size, crowding, and stimulus-driven changes in spatial attention.

The size of faces in the periphery may drive the differential VTC responses for Unscaled vs. Scaled faces. However, the effect of size may originate at different stages in the ventral processing stream. It is possible, for instance, that decreased visual acuity in the periphery which occurs in early visual stages (starting with the retina) may propagate forward and result in decreased VTC responses. In other words, perhaps the low response in face-selective cortex results from subjects' inability to see the faces? Several pieces of evidence suggest this is not the case. First, subjects reported being able to see faces presented in the periphery after scanning. Second, we collected behavioral data in a control experiment outside the scanner, where we showed similar face stimuli at a matched visual angle to the stimuli while scanning. For each of our three face mapping stimuli, subjects could classify the gender of faces in the bars well above chance, even in the periphery (mean performance > 70% for all eccentricity; significantly above chance, $p < 10^{-4}$, see Methods).

Another possibility is that size effects occur at higher stages of the visual processing hierarchy, at or near VTC. While detailed measurements of fMRI responses to different sized faces in the periphery have not yet been reported, electrophysiological recording in monkey IT cortex hint at this possibility. There is a tendency for larger RF size estimates studies that used larger mapping stimuli. For examples, (Tovee et al., 1994) showed monkeys large (8.5° and 17°) faces in the fovea and periphery, and found very modest position effects for IT neurons (RF sizes around 20°). By contrast, DiCarlo (2003) used small (0.6°) shapes and found high sensitivity to position (consistent with RF sizes around 2.5°).

If neural responses in VTC are sensitive to face size, this suggests it may serve as a neural correlate for reported behavioral effects of face size (Makela et al., 2001; Melmoth et al., 2000). These previous studies measured face recognition ability (discriminating a face among a set of similar faces) across varying sizes, eccentricities, and contrasts. One of the questions addressed was, can recognition performance for a foveal and peripheral face be matched by varying size and/or contrast? It was found that by increasing both size and contrast, a peripheral face could be recognized with the same accuracy as a foveal face. However, for faces of a given contrast, size adjustment alone is not sufficient to match foveal and peripheral face recognition. Because the face stimuli were presented at a constant contrast level, this result suggests that subjects would always be somewhat better at recognizing foveal than peripheral faces, regardless of the scaling employed. In this way, VTC responses may reflect a basic asymmetry in face recognition mechanisms between foveal and peripheral faces.

A second likely factor mediating the results is crowding: a decreased ability to recognize stimuli which are surrounded by other, nearby stimuli. Crowding is a significant bottleneck that affects our ability to see a wide range of objects (Pelli, 2008). Behavioral

data suggest that crowding is independent of object size, and increases when the stimuli are more visually similar (Pelli, 2008). Crowding, and stimulus context, are also known to modulate neural responses in monkey IT cortex. Estimated RF sizes for an object can be substantially reduced in the presence of a complex surrounding scene, if the object is part of a behaviorally relevant task (Rolls et al., 2003). The presence of multiple objects in the visual field can reduce neural responses in monkey IT cortex compared to isolated objects (Zoccolan et al., 2005).

Crowding may be determined by the separation of object representations along the cortical surface (Levi, 2008; Pelli, 2008). That is, the critical visual spacing between two stimuli corresponds to the minimal distance on the cortical surface to generate distinct responses. Some versions of this hypothesis suggest that crowding effects should derive from early visual cortex (Pelli, 2008). Other interpretations suggest that different crowding effects may occur at different stages of the processing hierarchy (Levi, 2008). Our results suggest that these hypotheses may be testable with methods similar to those employed here. While the conventional view of VTC as highly position-insensitive would suggest crowding effects could not occur there (since all of VTC would be expected to respond across visual space), our results indicate that small objects may produce measurably distinct response patterns across VTC for different positions, allowing estimates of the effect of surrounding stimuli on cortical patterns of response.

Finally, both the size and crowding of faces may affect VTC responses by producing shifts in the distribution of spatial attention. Different face stimuli may contain differing bottom-up salience, and consequently produce different exogenous shifts in spatial attention to different locations of the visual stimulus while subjects maintained fixation. That is, Full-Field faces may generate a large attentional spotlight extending across the entire extent of the face, while the Unscaled faces may facilitate a smaller, more focused attentional spotlight around the fovea, where the small faces are more salient. Physiological evidence shows that IT neurons' RF size can vary for tasks with different spatial attentional demands. (Richmond et al., 1983) showed that the presence or absence of a fixation point when monkeys made a judgment about simple stimuli strongly changed the measured RF size. Making the fixation point disappear right before the monkeys made the task judgment resulted in much broader RF sizes, whether the monkey was attending to the stimulus or fixation.

Do stimulus effects reflect spatial attention?

While an attentional explanation of our results is a distinct possibility, such an explanation must account for the profound regional-specificity of our results. We find strong stimulus-dependent position effects in VTC, but little to no such effect in V1 and other early visual regions. Previous measurements of spatial attentional affects on fMRI responses indicate robust effects across visual cortex, including in V1 (Gandhi et al.,

1999; Tootell et al., 1998) Furthermore, the magnitude of the effect varies markedly among different parts of VTC, with different face-selective regions exhibiting different degrees of sensitivity to stimulus content (**Figures 6-7, Table 1**). A global shift in spatial attention across stimuli would not be expected to produce the complex pattern of results we observe here. These regional differences may reflect different effects of top-down and bottom-up attentional mechanisms across cortex.

Are responses to Full-Field Faces sensitive to face features?

Full-Field faces stimulus may be different in a basic way from the Scaled and Unscaled stimuli. For this stimulus, only a single face was presented at a time, and the subject could only see a part of the face at any given moment in the experiment. One concern regarding the interpretation of positional effects with Full-field versus Scaled or Unscaled faces, is that for the full field faces, the appearance of specific facial features (e.g. eyes, mouth) is confounded with retinal position, because the position of the underlying face does not vary in this condition. In other words, do these effects reflect differential position sensitivity or differential responses to specific facial features?

Several aspects of the data indicate that the latter account is unlikely. First, the large pRF sizes indicate that each voxel would be responding to a range of local face features, rather than a particular well-localized feature. Second, such an account would predict that the pRF properties for Full-field faces should be qualitatively different than Scaled or Unscaled Faces, where face features occurred in all locations in the visual field. However, the pRF distributions for Full-field faces are significantly correlated with the pRF estimates from the Scaled and Unscaled stimuli—a similarity that is not predicted by a parts-based account. Finally, neither the distribution of pRF centers nor the pRF coverage maps for Full-field faces suggest that the pRFs preferentially cover particular salient parts of a face, such as the eyes, mouth or nose (**Figure 7)**.

Relation to eccentricity bias

Malach and colleagues (Hasson et al., 2003; Hasson et al., 2002; Levy et al., 2001; Levy et al., 2004; Malach et al., 2002) reported a coarse eccentricity bias across ventral temporal cortex. Category-selective regions located laterally along VTC, such as face-selective and word-form-selective regions, have preference for foveal stimuli over peripheral stimuli, while medial place-selective regions have a preference for peripheral stimuli over foveal stimuli. This organization is proposed to reflect different fixation demands for different stimulus classes: whereas word forms and faces are generally foveated upon, places may require integration of features across the visual field. These

eccentricity preferences may reflect an anatomical extension of iso-eccentricity lines extending from early visual cortex.

Several questions about the coarse eccentricity bias remain unresolved. One important question is whether retinotopic sensitivity is homogenous within each category-selective region. If the foveal bias in face regions originates from fixation demands of face recognition, this might suggest that all face-selective neural populations will uniformly prefer foveal stimulation. Our results indicate that face-selective regions exhibit a range of measurably-distinct position preferences (**Figure 6; Supplemental Figures 3-4**). Although the eccentricity distribution of the pRF estimates tends towards the fovea, many voxels preferring relatively more peripheral locations in visual space. For the IOG region in particular, these voxels extend well into the periphery, even for small stimuli. These across-voxel differences become more apparent when large face stimuli are used (**Figure 5**).

Might the eccentricity differences we observe reflect variations in face-selectivity across voxels? Different fMRI voxels may reflect different proportions of face-selective neural populations. It may be the case that face-selective neural responses are uniformly foveal, but these signals are mixed with other, less-foveally-biased signals in the fMRI responses. If this were the case, we would expect more highly face-selective voxels to exhibit a greater foveal bias: those measurements which more directly reflect face-selective neural responses should also have pRFs nearer the fovea. However, our analyses indicate that this is not the case (**Supplemental Figure 5**). We compared face-selectivity and eccentricity across the anatomically-defined extent of VTC, plotting the degree of face-selectivity measured during our localizer session against the eccentricity of the estimated pRF from the Full-Field face stimulus. For most subjects, there is no significant relationship between these two parameters.

Within face-selective ROIs, there is a basic asymmetry between foveal and peripheral voxels. Voxels with peripheral pRFs will cover the fovea, but foveal pRFs will not cover the periphery. Therefore, peripheral voxels respond strongly to foveal stimulation, while voxels with foveal pRFs exhibit weak to no responses when a face stimulus is in the periphery. As a consequence, the mean response across face-selective regions exhibits a stronger response to foveal than peripheral faces (**Figure 8**; (Hasson et al., 2002; Levy et al., 2001; Malach et al., 2002)). This strong foveal bias in the mean signal may mask the variability across voxels which we observe. Our results indicate that substantial information about stimulus position is present in the distributed response across voxels. This feature may be exploited by classifiers that can infer the position of stimuli from these patterns (Carlson et al., 2009)).

Regional differences in position effects across face-selective ROIs

Finally, there are qualitative and quantitative differences in retinotopic properties across three anatomically distinct face-selective regions. The most pronounced differences are between the IOG and mFus, while the pFus illustrates an intermediate profile, which more closely resembles mFus. For any given mapping stimulus, the pRFs have different distributions across IOG, pFus, and mFus. Compared to the more anterior pFus and mFus regions, the IOG region consistently has pRFs that extend further into the periphery (**Figure 6**). Moving from posterior to anterior across the ROIs, we observe a progressively stronger foveal bias, and greater coverage of the ipsilateral visual field (**Figure 7**). Consequently, the IOG illustrates significantly higher responses to contralateral versus ipsilateral stimulation as compared to the more anterior ROIs. These differences in coverage were confirmed in data from a separate experiment, where smaller faces were presented one at a time (**Figure 8**).

A second difference across face-selective regions is their sensitivity to stimulus content. The magnitude of the stimulus dependence effect also varies across the ROIs, with a relatively modest effect in IOG, a stronger effect in pFus, and a highly pronounced effect in mFus. MFus illustrates a larger increase in pRF size for larger faces, as compared to pFus or IOG. On the other hand, IOG shows a larger shift in the location of pRF centers compared to the mFus. IOG also exhibits a stronger correlation in the relative pattern of pRFs across voxels, perhaps because pRFs extend into the periphery for each of the three mapping stimuli for IOG but not pFus or mFus.

These data suggest a hierarchy across face-selective regions, with IOG, then pFus, then mFus placed at progressively higher stages in the hierarchy. As one ascends the hierarchy, coverage of visual space becomes less similar to a visual field map, extends further into the fovea, and covers more of the ipsilateral visual field. These properties are all consistent with early measurements of monkey IT cortex (Gross et al., 1969; Gross et al., 1972). However, we note that position effects at higher stages in the hierarchy, such as mFus, appear to be *more* sensitive to the stimulus, not less. This seems somewhat counterintuitive in the context of more traditional position-insensitive models of the ventral visual pathway, where the highest levels discard position information. We propose that face information in these regions is represented in a distributed manner, with both position and stimulus information playing a strong role in the population response. This view is consistent with recent theoretical (DiCarlo and Cox, 2007), behavioral (Afraz and Cavanagh, 2008; Dill and Edelman, 2001; Newell et al., 2005), and physiological (Zoccolan et al., 2007) studies which suggest that object recognition may exploit a combination of position and shape information.

These results also raise the question of whether other parts of extrastriate visual cortex may exhibit stimulus-dependent position effects. Face-selective regions have been reported lateral and dorsal to VTC, along the posterior superior temporal sulcus (Pinsk et al., 2009; Pinsk et al., 2005), while visual field maps have been reported nearby in temporal occipital cortex, around the hMT+ complex (Amano et al., 2009). It is not clear if the measured position effects in these regions have a similar stimulus sensitivity to VTC or not. These measurements may help reveal how the functional organization of dorsal face- and motion-selective regions relates to VTC, as well as whether dorsal face-selective regions are part of the same processing hierarchy as IOG, pFus, and mFus.

Conclusions

Retinal position plays a strong role in determining the measured fMRI response across ventral temporal cortex. The effect of position can be quantitatively interpreted using pRF methods, but the pRF approach only provides a partial description of the response properties in these regions. In this region of cortex, and likely in lateral occipital cortex, the time series is not adequately modeled by a static pRF.Rather, a full model of response will require a more complete model that accounts for the interaction between the effects of position and stimulus content.

Experimental Procedures

Subjects

Seven subjects participated in this study (4 female, ages 24-40). All subjects were right-handed and had normal or corrected-to-normal vision. Three subjects are authors of the paper. Each subject participated in 2 or more functional scanning sessions: a category localizer session, in which different object categories were presented for the purpose of localizing object-selective and category-selective regions; and a pRF mapping session, in which population receptive field mapping stimuli were shown. All experiments were undertaken with the written consent of each subject. Procedures were approved in advance by the Stanford Internal Review Board on Human Subjects Research.

Stimuli

For the category localizer session, stimuli were grayscale images of different object categories: faces, body parts (limbs), flowers, cars, guitars, guitars, and flowers. 12-second blocks of different object images were interspersed with blocks of two different baseline conditions: piecewise scrambled versions of the object images (drawn evenly from all object categories), and a "blank" condition in which a mean-luminance screen with a fixation point was presented for 12 seconds. (For subsequent analysis purposes, the blank period was considered the baseline, and the scrambled blocks analyzed in the same manner as each object category.) Subjects performed a one-back matching task during the localizer experiments.

For the main pRF experiments, stimuli comprised four different conditions of grayscale images masked by a moving bar aperture (**Figure 1a**). Each condition used frontal-view faces from an image database of ~100 high-resolution, normalized face images collected at Stanford University. Face stimuli had an equal proportion of male and female faces. The movement of the bar aperture was identical to the pRF-mapping experiments described in (Amano et al., 2009; Dumoulin and Wandell, 2008).

The four stimulus conditions in the main experiment were as follows. **(a)** "Unscaled Faces": faces subtended ~2° in diameter uniformly across the visual field, spaced closely together (~0.5° of gray space between faces). The faces were arranged with a single face at the fovea, and rings of faces surrounding the central face in different bands of equal eccentricity. Every two seconds, the field would cycle through six randomly-generated face fields, each viewed through the same bar aperture. The position of faces in each eccentricity band rotated around the center, to show different face parts within the aperture. Different faces were randomly assigned to each position for each face field, so that over the course of scanning many faces were presented at

each visual location. Over the course of the experiment, 24 different face fields were shown. **(b)** "Scaled Faces": three different eccentricity bands of faces were presented (centered at 0, 3.5, and 10°), with faces in each band scaled to twice the diameter of the previous band (2, 4, and 8° diameters). The spacing between face edges was 0.5°, as with the Unscaled condition, although because of the size scaling, the distance between face centers increased (3.5 and 6.5° between successive eccentricity bands). The eccentricity bands rotated, and face identities changed, in the same fashion as the Unscaled condition. **(c)** "Full-field Faces": a single face was presented one at a time, centered on the fovea and spanning the entire ~28° diameter of the display. Every two seconds, the display cycled through six randomly-selected faces, with the face parts (eyes, nose, mouth, chin) of each face presented at the same location in visual space. **(d):** "Phase-Scrambled Faces": the stimuli from the "Scaled Faces" condition were converted into frequency space, the phase component was randomly reordered while the spatial frequency component was kept constant, and the image was reconstructed from frequency space with an inverse Fourier transform. The resulting images had the same spatial frequency content as the intact faces, but resembled cloudlike textures (Honey et al., 2008).

Subjects participated in at least one data collection session, and 4 experimental runs of each condition. (The total number of runs per condition varied from 4 to 12 across subjects.) For each experimental run, the individual faces used to generate the stimuli were randomly selected. Each subject had seen all of the face images used many times over the course of the experiments, and so were familiarized with each image. Images were randomly repeated so that no individual face image was correlated with any particular part of visual space more than any other.

In the scanner, stimuli were projected on to a screen mounted on the cradle for the imaging coil. Subjects lay on their backs in the bore of the MR scanner and viewed the screen through an angled first-surface mirror positioned approximately 8 cm in front of their eyes. The mean distance between subjects' eyes and the stimulus image was 22 cm; the maximum size subtended by a stimulus at this distance was 28.5° (14.25° radius from the fovea).

Behavioral Control Experiment

In order to ensure consistent fixation, subjects were asked to perform a color-discrimination task while scanning. The fixation disc would briefly flash one of three possible colors, and subjects responded with the color. Although subjects were fixating, they reported being able to see faces across the different mapping conditions, with the exception of Phase-scrambled Faces.

We assessed how well subjects could see the faces across the three mapping stimuli with a separate behavioral control experiment outside the scanner. In this behavioral experiment, subjects viewed vertical bar apertures containing the same three face conditions as in the scanner, except that each stimulus contained images from a single gender that occurred with a 50% probability. The stimuli were presented with the same visual angle as the stimuli in the scanner, in a darkened psychophysics room on a contrast-calibrated Hewlett-Packard HPLP2480zx display.

Subjects completed a series of six runs of 120 trials each. During each trial, a bar stimulus was flashed briefly for 150 ms at a random position while subjects maintained fixation. Each bar contained either only male, or only female faces, with a 50% probability of each. Subjects classified the faces as male or female.

During the behavioral experiment, subjects consistently performed above chance for all bar positions and all stimuli (all p values $< 10^{-4}$ for all positions, one-tailed T test across subjects). Performance was consistently better for more foveal positions, being near ceiling (up to 98% correct for the central bar positions) and decreasing with eccentricity (two-way ANOVA with factors of stimulus type and bar eccentricity: stimulus type: $F(2,15) = 3.66$, $p = 0.03$; eccentricity: $F(5,20) = 9.94$, $p < 10^{-30}$; interaction: $F(17,4) = 1.44$, $p = 0.17$). But performance remained above 72% correct for all stimuli and eccentricities (**Table 3**).

Eccentricity (°)	Unscaled	Scaled	Full-Field
1	96.1 (1.0)	95.4 (1.2)	87.9 (2.2)
3.1	94.6 (1.5)	93.9 (1.8)	91.8 (1.8)
5.1	92.1 (1.1)	95.0 (2.5)	90.7 (2.3)
7.2	87.5 (3.1)	85.4 (3.8)	90.4 (2.5)
9.3	92.1 (1.7)	86.4 (3.1)	90.0 (1.2)
11.3	82.1 (3.5)	72.1 (6.7)	73.6 (5.4)

Table 3. Performance on the behavioral control task. Each row indicates the mean distance of the bar aperture containing faces from the fixation mark. Apertures to the left and right of fixation are averaged together. Each entry indicates the mean percent correct across subjects, with the standard error of the mean indicated in parentheses (N = 7 subjects).

MR Data Acquisition

All MR data were acquired at the Stanford Lucas Imaging Center using a 3T GE Signa scanner. We utilized a custom receive-only occipital quadrature RF surface coil (Nova Medical, Inc., Wilmington, MA, USA). The dimensions of the coil are: interior dimension

left/right = 9 inches, exterior dimension left/right = 10 1/8 inches; height = 5 1/4 inches; length = 7.5 inches.

Functional data from the main pRF-mapping experiments were collected at 1.7 x 1.7 x 2.5 mm resolution (with an effective resolution of 2.5 mm isotropic). We collected 20 slices oriented perpendicular to the calcarine sulcus covering occipito-temporal cortex of both cortical hemispheres and the following parameters: TE = 30 ms, TR = 1500 ms, single-shot, Flip angle = 71°, FOV = 220 mm. During each session we acquired inplane anatomical images at the same prescription as the functional scans. Anatomical inplanes were used to co-register each session's data to the subject's whole brain anatomy.

Data for the localizer experiment were collected during a separate scanning session from the main experiments. We collected 12 oblique slices around extrastriate cortex, using voxels of size 1.5 x 1.5 x 3mm. For one subject (S1), we also used a second localizer scan, with the same imaging parameters, but a voxel size of 1.5 x 1.5 x 1.5mm. Functional scans used the following parameters: TE = 30 ms, TR = 1000 ms, two-shot, Flip angle = 77°, FOV = 192 mm. This localizer data was also used for a separate study described in (Weiner and Grill-Spector, 2009).

Anatomical Data Analysis

For each subject, we obtained a high-quality whole brain anatomical volume during a separate scanning session. These anatomies were collected using a 1.5 T GE Signa scanner and a head coil at the Lucas Imaging Center. Data across different functional imaging sessions were aligned to this common reference anatomy, in order to use the same ROIs across sessions, and restrict activations to gray matter. The whole brain anatomy was the average of three T1-weighted SPGR scans (TR = 1000 ms, FA = 45°, 2 NEX, FOV = 200 mm, inplane resolution = 0.78 x 0.78 mm, slice thickness 1.2 mm, sagittal prescription) which were aligned using the mutual-information coregistration algorithm in SPM (Maes et al., 1997) (http://www.fil.ion.ucl.ac.uk/spm/). The reference anatomy was then segmented into gray and white matter using ITKGray and mrGray (http://white.stanford.edu/software/). The resulting segmentations were used to restrict activity to gray matter, and produce a mesh representation of the gray matter surface (Teo et al., 1997) for visualization of activations on the cortical surface.

fMRI Data Analysis

fMRI data were analyzed in MATLAB using the Stanford mrVista toolbox (http://white.stanford.edu/newlm/). BOLD data were temporally interpolated to adjust for slice-acquisition timing, and motion corrected via a robust affine transformation of each

temporal volume in a data session to the first volume of the first scan (Nestares and Heeger, 2000). This correction produced a motion estimate for each scan; scans with greater than 1.5 mm estimated net motion were removed from the analysis. Data were temporally high-pass filtered with a 1/40 Hz cutoff. We converted the time courses into percent signal change by dividing each temporal volume by the mean signal for that voxel.

ROI Selection

All regions of interest (ROIs) were defined off the segmented gray matter surface. ROIs were restricted to avoid regions of low signal intensity, which may indicate sinus artifacts (Winawer et al., 2010). Each ROI was transformed into the anatomical volume, where subsequent analyses on the data were performed. ROIs were inspected in the volume to ensure they remained anatomically contiguous and did not jump across sulcal banks, or extend into CSF.

ROIs defined from visual field maps were identified by reversals in the representation of polar angle, as described by (Wandell et al., 2007). This ROI selection was performed on separate retinotopic mapping sessions, which utilized a rotating wedge design and a traveling wave analysis. These experiments are described in detail in (Sayres and Grill-Spector, 2008).

Face-selective ROIs were selected based on the results of a general linear model (GLM) applied to the category localizer data (Burock and Dale, 2000). Because the functional data for the localizer session was slightly higher resolution than the main runs (1.5mm vs. 2.5mm in-plane resolution), and the higher resolution produced slightly patchy ROIs on the cortical surface, we selected the face-selective ROIs on functional data that had been spatially smoothed with a 2mm Gaussian kernel. Statistical contrast maps were then computed based on a T test of faces greater than limbs, flowers, cars, guitars, and houses. Each ROI was defined as a contiguous patch along the cortical surface with a significance level of $p < 0.001$ (uncorrected) for the given contrast. For subsequent analyses of the localizer data (the correlations between face-selectivity and pRF size and eccentricity, **Supplemental Figure 5**), we used the unsmoothed data.

Three separate face-selective ROIs were defined: (1) IOG: this ROI lies on or near the inferior occipital gyrus, and is a subset of region LO, the lateral and posterior section of the lateral occipital complex. (2) pFus: this ROI lies along the posterior end of the fusiform gyrus and occipitotemporal sulcus, lateral and slightly anterior to hV4. (3) mFus: this region lies along the middle of the fusiform gyrus, anterior to pFus. The anatomical delineation of these regions is consistent with that reported in (Pinsk et al., 2009) and (Weiner and Grill-Spector, 2009). These regions are separated by other regions which

are selective to body parts (Weiner and Grill-Spector, 2009). Because the more commonly-used designations "FFA" and "OFA" can be ambiguous with respect to which of the three ROIs they refer, for this paper we will use these anatomically-referenced designations.

In addition to visual field maps and face-selective regions, we also defined an anatomical ROI that encompassed the ventral surface of temporal cortex (VTC). We excluded the face-selective regions from the VTC ROI, This ROI is used to test whether our observed effects occur only in face-selective regions, or more broadly across ventral cortex; and to measure larger-scale variations in retinotopic sensitivity, such as the medial/lateral gradient between foveal/peripheral representation. The VTC ROI was manually drawn in each hemisphere to extend from the collateral sulcus at the medial boundary, to the inferior temporal gyrus at the lateral boundary. It was also selected to exclude known visual field maps hV4 and VO-1 (posterior/medial boundaries), as well as signal falloff from the ear canal (anterior boundary).

Ventral temporal cortex lies near the transverse sinus, a blood vessel which is known to produce signal distortions in nearby anatomical regions (Jezzard and Ramsey, 2003; Winawer et al., 2010). The precise extent and magnitude of these distortions is not fully understood. However, for spiral gradient echo scans such as the ones we collected for our experiments, many of the regions with the most serious distortions are indicated by a low mean intensity in the raw functional images (Winawer et al., 2010). Therefore, we excluded from all our analyses voxels with a mean image intensity of less than 40% of the maximum intensity within gray matter.

Population Receptive Field (pRF) Model

The main experiments were analyzed using a population receptive field (pRF) model described in detail in (Dumoulin and Wandell, 2008). Each pRF is represented as a two-dimensional Gaussian function in visual space, described by the center (x_0, y_0)—or (r, θ) when expressed in polar coordinates—and standard deviation σ. A given pRF can be used to predict the time course of response for a given stimulus, in this case the movement of the bar stimuli across the visual field. This prediction is computed by multiplying the pRF with a binary mask of the stimulus position at each time point, and convolving the resulting series of amplitudes over time with a hemodynamic impulse response function (HIRF). We used the SPM difference-of-gammas HIRF for our analyses (Friston et al., 1998).

The pRF analysis solves for each voxel the set of parameters (r, θ, σ) which best fits the observed response for that voxel. The solution process proceeded in a two stage,

coarse-to-fine approach. In the first stage ("grid fit"), a triangulated grid of ~100,000 pre-set parameter values (r, θ, σ) were used to produce a set of predicted responses. These predictions were fit to the responses of each voxel, and the best-fitting set of parameters was selected. In the second stage ("search fit"), the best parameters from the grid fit were used as a starting point for a constrained-optimization search to find the final (r, θ, σ) for each voxel.

The fitting process produced the parameter estimates r (eccentricity), θ (polar angle) and σ (size) for each voxel, as well as a goodness-of-fit parameter, R^2 (the proportion of the time courses' variance explained by the best-fitting model). For most analyses, we used the R^2 metric to restrict our analyses to voxels which were well-fit by the pRF model and eliminate noisy voxels. We restricted our analyses to voxels which had 15% variance explained or greater by the model. We also converted the size parameter from a standard deviation σ to a half-width half-maximum (HWHM) using the formula:

HWHM = sqrt(2 * ln(2)) * σ

We use this measure because it has the property that when the HWHM and eccentricity match, the response to foveal stimulation is exactly half the maximal response.

Position-Invariant Model

An alternate model to the pRF model is a position-invariant model, in which fMRI voxels are insensitive to the retinal position of a stimulus, but respond to the presence or absence of the stimulus. Because the stimulus had blank periods in which no faces were present, even a position-insensitive model predicts some modulation of responses. We modeled the predictions of this response by creating a predictor whose value was 1 when a face was present, and 0 when a face was absent, and convolving this predictor with the SPM HIRF (same as for the pRF model). We applied the prediction of this model to all voxels in VTC, to estimate the goodness of fit of this alternate model compared to the pRF model. The results are illustrated in **Figure 2** and summarized in **Supplemental Figure 1.**

pRF Coverage Measurements

We measured the coverage of the visual field for each ROI in each subject. This coverage defines the extent of visual space for which a stimulus effectively elicits a response from the ROI. We followed the procedures described in (Amano et al., 2009; Winawer et al., 2010). To produce a coverage map for a given ROI, we took the 2D pRF in visual space from all voxels with at least 15% variance explained by the model. Each pRF was normalized to range from 0 to 1. For each point in visual space, we took the

maximum value across all pRFs. In order to reduce the effect of outlier voxels on coverage levels, we smoothed the size parameter using a median filter. For each voxel, we found the two nearest neighbors in stimulus space (the two voxels with the closest pRF centers), and replaced that voxel's size parameter with the median of the three voxels' sizes. We further reduced the effect of outliers, and weighted the coverage by the density of pRFs in visual space, through 100 iterations of a bootstrapping procedure. In each iteration, we randomly sampled 80% of the above-threshold voxels in the ROI with replacement, and generated a single coverage map. We then took the mean of all 100 coverage maps as the final coverage map for that ROI and subject. Finally, we took the average coverage map across all subjects for the maps shown in **Figure 7**. We also present the distribution of the pRF centers across visual space, without taking into account the pRF size, in **Supplemental Figure 4**.

Block-design Face-Position Experiment

We tested predictions of the pRF coverage analyses (**Figure 7**) by measuring responses to faces presented at different retinotopic positions in a block design (**Figure 8**). We used data from a 6-position experiment which has previously been described in (Sayres and Grill-Spector, 2008). In this experiment, grayscale images of faces were presented in 12-second blocks, with each face appearing at a rate of 1 Hz, alternating with 12-s mean-luminance fixation (baseline) blocks. Face images subtended 2.5° visual angle, and were presented at six positions: centered at the fixation point (foveal); centered 4° above, below, to the left, and to the right of fixation; and centered 9° to the left of fixation. Of these six positions, we analyzed three positions for the present study: foveal, 4° ipsilateral to each hemisphere, and 4° contralateral to each hemisphere. These experiments also included blocks of body-part and house images; for the purpose of the present study, we focused on measuring the response to face images, because pRFs for other stimuli may differ.

We analyzed data from 5 subjects who participated in both the face pRF experiments, and the block-design position experiment. We extracted responses from each of the face-selective ROIs (IOG, pFus, mFus) for each hemisphere. We summarized results across both hemispheres by sorting conditions into ipsilateral, contralateral or foveal stimulation (e.g., faces presented in the right visual field would be considered ipsilateral to a right-hemisphere ROI, and contralateral to a left-hemisphere ROI). Mean responses were averaged across hemispheres for each condition. We measured the mean peri-stimulus time course for each position, and the mean beta coefficient from a general linear model. We computed the GLMs in each hemisphere, using the same difference-of-gammas HIRF as for the pRF estimates, and previously described methods (Sayres and Grill-Spector, 2008).

Acknowledgements

We thank Aviv Mezer and Jonathan Winawer for fruitful discussions. This study was funded by NIH grant NEI EY03164 to BW, NSF BCS-0617688 to KGS and NIH 1R21EY017741 to KGS.

References

Afraz, S. R., and Cavanagh, P. (2008). Retinotopy of the face aftereffect. Vision Res *48*, 42-54.

Amano, K., Wandell, B. A., and Dumoulin, S. O. (2009). Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. J Neurophysiol.

Boussaoud, D., Desimone, R., and Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. J Comp Neurol *306*, 554-575.

Burock, M. A., and Dale, A. M. (2000). Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. Hum Brain Mapp *11*, 249-260.

Carlson, T., Hogendoorn, H., Fonteijn, H., and Verstraten, F. A. (2009). Spatial coding and invariance in object-selective cortex. Cortex.

DiCarlo, J. J., and Cox, D. D. (2007). Untangling invariant object recognition. Trends Cogn Sci *11*, 333-341.

DiCarlo, J. J., and Maunsell, J. H. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. J Neurophysiol *89*, 3264-3278.

Dill, M., and Edelman, S. (2001). Imperfect invariance to object translation in the discrimination of complex shapes. Perception *30*, 707-724.

Dumoulin, S. O., and Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. Neuroimage *39*, 647-660.

Ellis, R., Allport, D. A., Humphreys, G. W., and Collis, J. (1989). Varieties of object constancy. Q J Exp Psychol A *41*, 775-796.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. Neuroimage *7*, 30-40.

Gandhi, S. P., Heeger, D. J., and Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. Proc Natl Acad Sci U S A *96*, 3314-3319.

Goodale, M. A., and Humphrey, G. K. (1998). The objects of action and perception. Cognition *67*, 181-207.

Grill-Spector, K., Knouf, N., and Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci *7*, 555-562.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187-203.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., and Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum Brain Mapp *6*, 316-328.

Gross, C. G., Bender, D. B., and Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science *166*, 1303-1306.

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophysiol *35*, 96-111.

Hasson, U., Harel, M., Levy, I., and Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. Neuron *37*, 1027-1041.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., and Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. Neuron *34*, 479-490.

Honey, C., Kirchner, H., and VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. J Vis *8*, 9 1-13.

Hubel, D. (1995). Eye, Brain and Vision, Vol 22 (New York, N.Y., Scientific American Library / Scientific American Books).

Ito, M., Tamura, H., Fujita, I., and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol *73*, 218-226.

Jezzard, P., and Ramsey, N. F. (2003). Measuring changes caused by disease. In Quantitative MRI of the Brain, P. Tofts, ed. (Chichester, West Sussex

Hoboken, NJ, Wiley), pp. 413-454.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci *17*, 4302-4311.

Konen, C. S., and Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. Nat Neurosci *11*, 224-231.

Levi, D. M. (2008). Crowding--an essential bottleneck for object recognition: a mini-review. Vision Res *48*, 635-654.

Levy, I., Hasson, U., Avidan, G., Hendler, T., and Malach, R. (2001). Center-periphery organization of human object areas. Nat Neurosci *4*, 533-539.

Levy, I., Hasson, U., Harel, M., and Malach, R. (2004). Functional analysis of the periphery effect in human building related areas. Hum Brain Mapp *22*, 15-26.

Li, N., Cox, D. D., Zoccolan, D., and DiCarlo, J. J. (2009). What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? J Neurophysiol *102*, 360-376.

Lueschow, A., Miller, E. K., and Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. Cereb Cortex *4*, 523-531.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging *16*, 187-198.

Makela, P., Nasanen, R., Rovamo, J., and Melmoth, D. (2001). Identification of facial images in peripheral vision. Vision Res *41*, 599-610.

Malach, R., Levy, I., and Hasson, U. (2002). The topography of high-order human object areas. Trends Cogn Sci *6*, 176-184.

Melmoth, D. R., Kukkonen, H. T., Makela, P. K., and Rovamo, J. M. (2000). The effect of contrast and size scaling on face perception in foveal and extrafoveal vision. Invest Ophthalmol Vis Sci *41*, 2811-2819.

Milner, A. D., and Goodale, M. A. (1993). Visual pathways to perception and action. Prog Brain Res *95*, 317-337.

Nestares, O., and Heeger, D. J. (2000). Robust multiresolution alignment of MRI brain volumes. Magn Reson Med *43*, 705-715.

Newell, F. N., Sheppard, D. M., Edelman, S., and Shapiro, K. L. (2005). The interaction of shape- and location-based priming in object categorisation: evidence for a hybrid "what + where" representation stage. Vision Res *45*, 2065-2080.

Niemeier, M., Goltz, H. C., Kuchinad, A., Tweed, D. B., and Vilis, T. (2005). A contralateral preference in the lateral occipital area: sensory and attentional mechanisms. Cereb Cortex *15*, 325-331.

Op De Beeck, H., and Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. J Comp Neurol *426*, 505-518.

Pelli, D. G. (2008). Crowding: a cortical constraint on object recognition. Curr Opin Neurobiol *18*, 445-451.

Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., and Kastner, S. (2009). Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. J Neurophysiol *101*, 2581-2600.

Pinsk, M. A., DeSimone, K., Moore, T., Gross, C. G., and 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.

Puce, A., Allison, T., Gore, J. C., and McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. J Neurophysiol *74*, 1192-1199.

Richmond, B. J., Wurtz, R. H., and Sato, T. (1983). Visual responses of inferior temporal neurons in awake rhesus monkey. J Neurophysiol *50*, 1415-1432.

Rolls, E. T., Aggelopoulos, N. C., and Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. J Neurosci *23*, 339-348.

Sayres, R., and Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. J Neurophysiol *100*, 249-267.

Schwarzlose, R. F., Swisher, J. D., Dang, S., and Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. Proc Natl Acad Sci U S A *105*, 4447-4452.

Teo, P. C., Sapiro, G., and Wandell, B. A. (1997). Creating connected representations of cortical gray matter for functional MRI visualization. IEEE Trans Med Imaging *16*, 852-863.

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., and Bulthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. Eur J Neurosci *14*, 869-876.

Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., and Dale, A. M. (1998). The retinotopy of visual spatial attention. Neuron *21*, 1409-1422.

Tovee, M. J., Rolls, E. T., and Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. J Neurophysiol *72*, 1049-1060.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. Science *311*, 670-674.

Tsao, D. Y., Moeller, S., and Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. Proc Natl Acad Sci U S A *105*, 19514-19519.

Ungerleider, L. G., and Haxby, J. V. (1994). 'What' and 'where' in the human brain. Curr Opin Neurobiol *4*, 157-165.

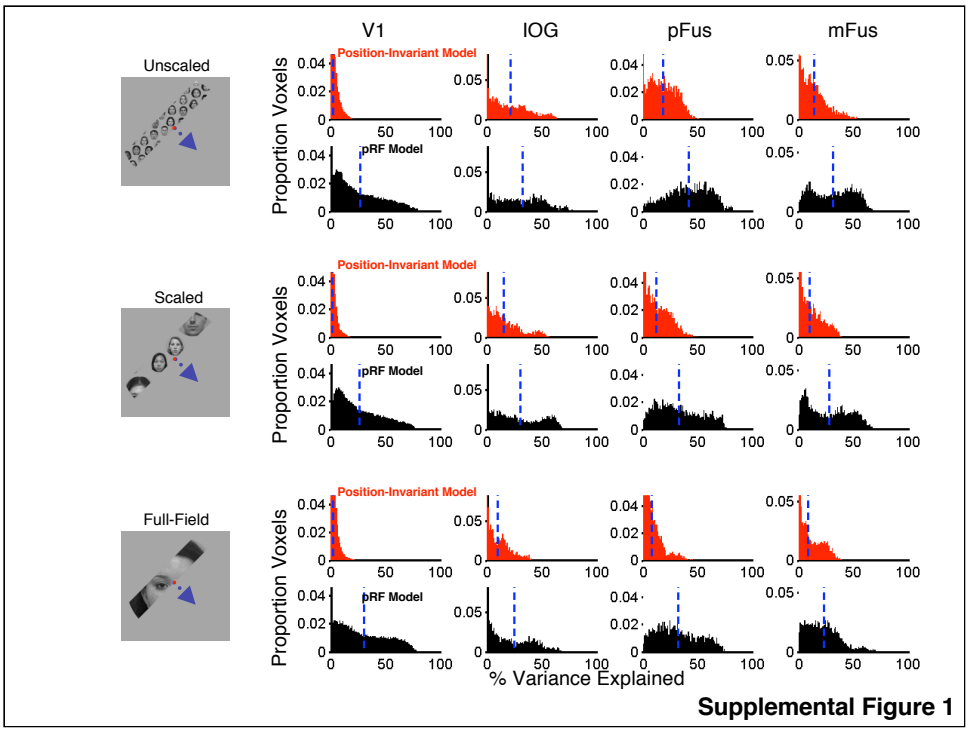
Wandell, B. A., Dumoulin, S. O., and Brewer, A. A. (2007). Visual field maps in human cortex. Neuron *56*, 366-383.

Weiner, K. S., and Grill-Spector, K. (2009). High-resolution fMRI reveals sparsely distributed representations of faces and limbs in human ventral temporal cortex. Paper presented at: Society for Neuroscience Annual Meeting (Chicago).

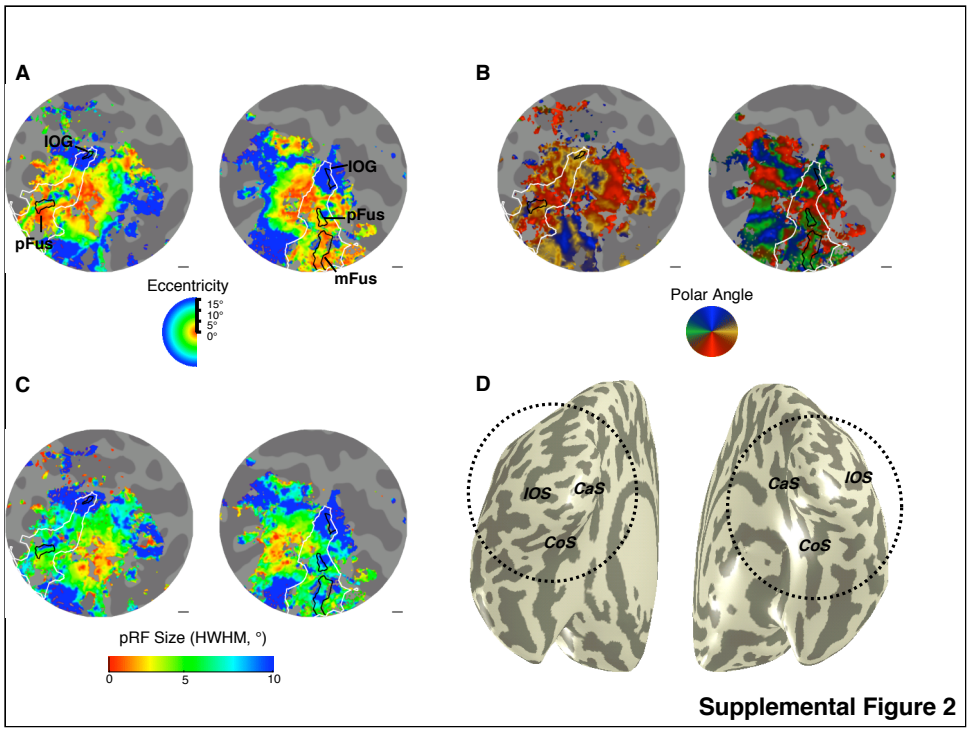
Winawer, J., Horiguchi, H., Sayres, R., Amano, K., and Wandell, B. (2010). Mapping hV4 and ventral occipital cortex: The venous eclipse. Vision Res *in press*.

Zoccolan, D., Cox, D. D., and DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. J Neurosci *25*, 8150-8164.

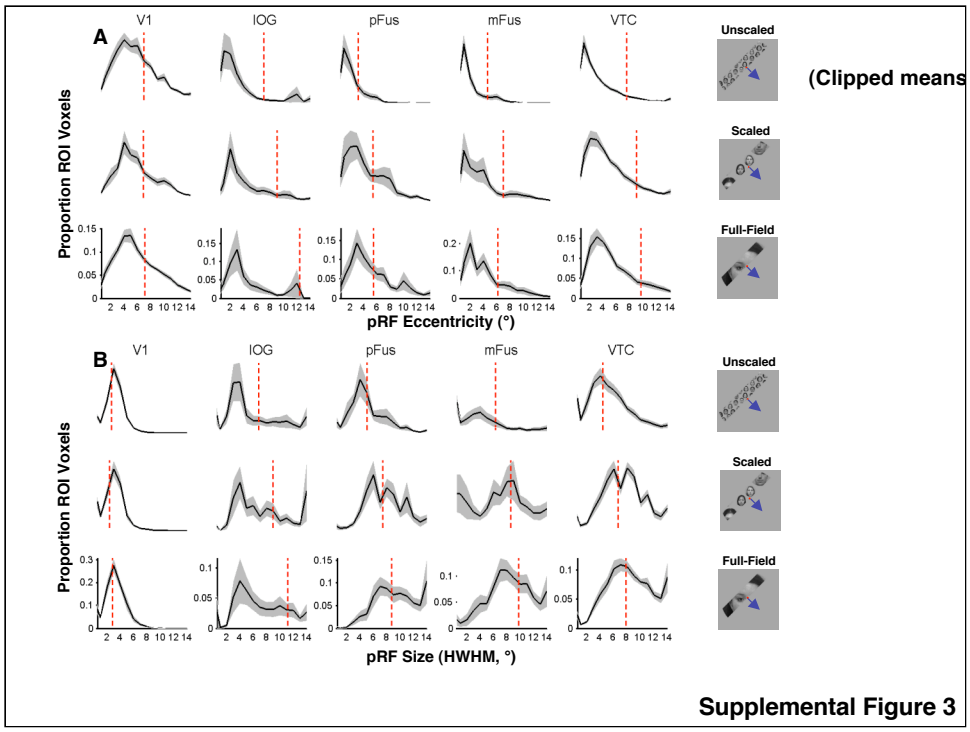
Zoccolan, D., Kouh, M., Poggio, T., and DiCarlo, J. J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. J Neurosci *27*, 12292-12307.



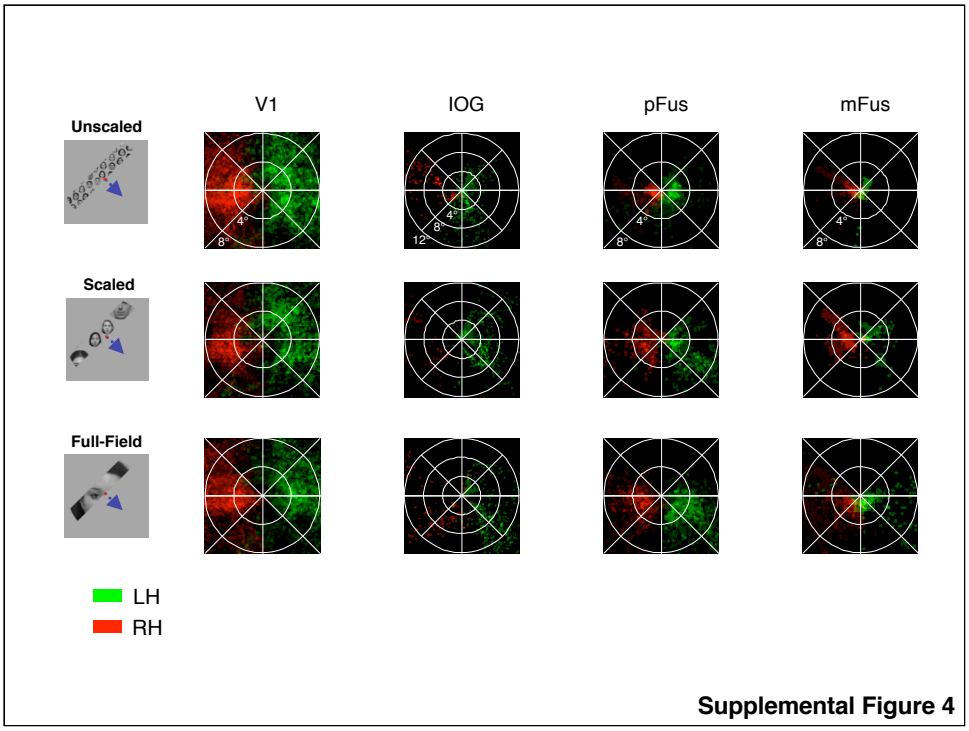
Supplemental Figure 1. Comparing the pRF model to a position-invariant model of fMRI responses. Because our stimuli included blank periods with no face stimulus, we would expect BOLD modulations even if a voxel was not sensitive to retinal position. We modeled a hypothetical voxel which responded to faces versus no faces, but did not respond to the face position, by constructing a design matrix indicating when faces were and were not present. We convolved this design matrix with the same hemodynamic impulse response function that we used in our pRF model. These histograms show the distribution of variance explained by this position-invariant model, compared to the position-sensitive pRF model, for each stimulus (*rows*) and region of interest (*columns*). The upper red histograms in each pair show the variance explained by the position-invariant model, and the lower black histograms show for the pRF model. The vertical blue dotted lines in each histogram show the mean variance explained. For each stimulus and ROI, the pRF model explained a significantly greater amount of response variance than the position-invariant model (Kolgomorov-Smirnov test: $p < 10^{-30}$, all comparisons, Bonferroni corrected).



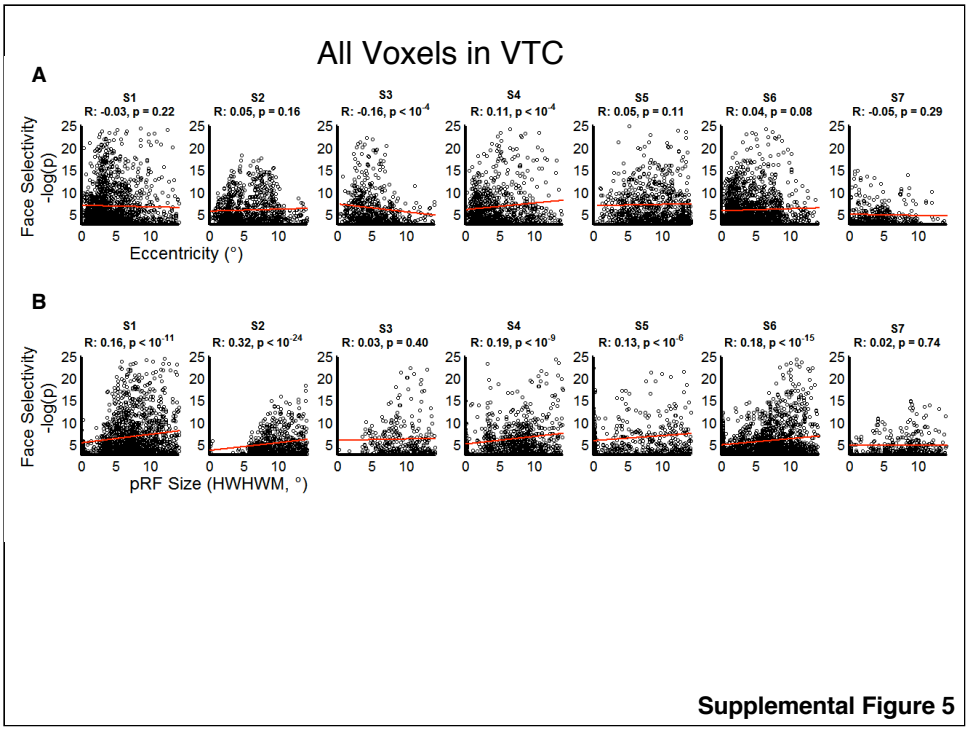
Supplemental Figure 2. Flat maps of the pRF parameters for the Scaled Faces stimulus. These maps show the pRF estimates in early visual cortex as well as lateral occipital and ventral temporal cortex of subject S2. **(a)** Eccentricity maps for the left and right hemispheres, respectively. The white outline indicates the anatomical VTC definition; the three face-selective regions (IOG, pFus, and mFus) are outlined in black and labeled. The scale bars indicate 1cm. **(b)** Polar angle maps. **(c)** Maps of pRF size (half-width-half-max). **(d)** Posterior view of the inflated cortical surface, indicating the location of the flat maps. Sulci are labeled in each hemisphere: CaS = calcarine sulcus; IOS = inferior occipial sulcus; CoS = collateral sulcus.



Supplemental Figure 3. Distributions of estimated pRF eccentricity and size for each mapping stimulus and ROI.



Supplemental Figure 4. Distribution of pRF centers for each ROI and mapping stimulus. Each plot shows the density of pRF centers, ignoring the pRF size estimates, for a given face stimulus (different rows) and ROI (different columns). Left hemisphere voxels for each ROI are colored green; right-hemisphere voxels are colored red. The brightness of each pixel reflects the number of voxels with a pRF at that visual position. These maps demonstrate the hemipsheric lateralization of visual field coverage for each ROI, as well as the differing eccentricities for pRF centers across ROIs and stimuli.



Supplemental Figure 5. Relationship between estimated pRF parameters and face-selectivity in VTC. These scatterplots plot the estimated pRF eccentricity **(a)** and size **(b)** for the Scaled Faces stimulus, against the estimated face-selectivity from the separate localizer session. Data come from across the anatomical VTC ROI, including face-selective voxels. Each column represents data from a different subject. Within each scatterplot, each circle represents one voxel. Face-selectivity is measured as the significance level (-log(p value)) from a T test of Faces > Inanimate objects. The red line shows the fit from a linear regression.