Eye-centered encoding of visual space in scene-selective regions

Emily J. Ward
Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania, PA, USA

Sean P. MacEvoy
Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania, PA, USA

Russell A. Epstein
Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania, PA, USA

We used functional magnetic resonance imaging (fMRI) to investigate the reference frames used to encode visual information in scene-responsive cortical regions. At early levels of the cortical visual hierarchy, neurons possess spatially selective receptive fields (RFs) that are yoked to specific locations on the retina. In lieu of this eye-centered organization, we speculated that visual areas implicated in scene processing, such as the parahippocampal place area (PPA), the retrosplenial complex (RSC), and transverse occipital sulcus (TOS) might instead possess RFs defined in head-, body-, or world-centered reference frames. To test this, we scanned subjects while they viewed objects and scenes presented at four screen locations while they maintained fixation at one of three possible gaze positions. We then examined response profiles as a function of either fixation-referenced or screen-referenced position. Contrary to our prediction, the PPA and TOS exhibited position–response curves that moved with the fixation point rather than being anchored to the screen, a pattern indicative of eye-centered encoding. RSC, on the other hand, did not exhibit a position–response curve in either reference frame. By showing an important commonality between the PPA/TOS and other visually responsive regions, the results emphasize the critical involvement of these regions in the visual analysis of scenes.

Keywords: functional imaging, space and scene perception, reference frame, receptive fields, parahippocampal place area

Citation: Ward, E. J., MacEvoy, S. P., & Epstein, R. A. (2010). Eye-centered encoding of visual space in scene-selective regions. *Journal of Vision, 10*(14):6, 1–12, http://www.journalofvision.org/content/10/14/6, doi:10.1167/10.14.6.

Introduction

A core concern of visual neuroscience is the identification of spatial reference frames used for visual processing. Visual information is initially obtained by the retina, which reencodes the visual world with every eye movement. Retinotopic organization is maintained in the lateral geniculate nucleus, primary visual cortex, and throughout a variety of extrastriate areas. Although one might suppose that higher level ventral stream regions might take advantage of alternate reference frames, such as an allocentric coordinate system tied to the principal axis of each object (Marr & Nishihara, 1978), or a world-centered reference frame tied to the fixed surfaces of the environment, data from neuroimaging studies tend to argue against this idea. Even high-level object processing regions such as the lateral occipital complex (LOC; Grill-Spector et al., 1999; Kourtzi & Kanwisher, 2001; Malach et al., 1995) exhibit position-dependent responses (Sayres & Grill-Spector, 2008; Wandell, Dumoulin, & Brewer, 2007), which are often assumed to be retinotopic. Indeed, a recent study of 12 visual areas, including V1–V4, MT, and several subregions of LOC, found that all 12 regions exhibited retinotopic responses (Gardner, Merriam, Movshon, & Heeger, 2008). Notably, this study dissociated retinotopic (eye-centered) from spatiotopic (screen-centered) responses by demonstrating that the response maps in each region shifted when the fixation point was displaced along the screen. Thus, both early visual areas and high-level object processing regions appear to code information in eye-centered coordinates.

Here we address the reference frame question for scene-selective regions, which were not examined in previous studies. Three regions of the human cortex respond more strongly when subjects view scenes such as landscapes, cityscapes, or rooms than when they view individual discrete objects such as faces, tools, vehicles, or appliances: the parahippocampal place area (PPA; Aguirre, Zarahn, & D'Esposito, 1998; Epstein, Harris, Stanley, & Kanwisher,

doi: 10.1167/10.14.6 Received May 25, 2010; published December 3, 2010 ISSN 1534-7362

1999; Epstein & Kanwisher, 1998; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999), the retrosplenial cortex/parietal–occipital sulcus region (RSC; Maguire, 2001), and the transverse occipital sulcus (TOS; Grill-Spector, 2003; Hasson, Harel, Levy, & Malach, 2003; Nakamura et al., 2000). Neuropsychological and neuroimaging studies suggest that these scene-responsive regions mediate place recognition and other functions that are critical to our ability to navigate accurately through the world (Aguirre & D'Esposito, 1999; Epstein, 2005; Maguire et al., 1998; Mendez & Cherrier, 2003). Because navigation may require world-centered or body-centered coordinates, one might hypothesize that these areas could encode visual information in a non-retinotopic reference frame. Indeed, one might speculate that the category preferences in scene-selective (PPA, RSC, TOS) vs. object-selective (LOC) regions might reflect the use of different reference frames for navigation vs. object recognition, as images of scenes and buildings convey information about the location and orientation of the viewer relative to fixed-to-the-earth scene elements, whereas images of isolated objects do not.

Previous work has shown that the PPA, TOS, and to a certain extent RSC are sensitive to the positions of visual stimuli (Arcaro, McMains, Singer, & Kastner, 2009; MacEvoy & Epstein, 2007). However, no study has yet sought to determine the reference frame within which this sensitivity is expressed. In previous studies, subjects fixated on a single screen location that was fixed relative to the head, body, and world leaving a potential head-, body-, or world-centered reference frame indistinguishable from an eye-centered one. In the present study, we disambiguated these reference frames by presenting visual stimuli at different screen locations while gaze position was varied. We predicted that response curves would be anchored to the fixation point in regions that encode information in eye-centered (retinotopic) coordinates but anchored to the screen in regions that encode information in head-, body-, or screen-centered coordinates. Stimuli were either scenes or objects, which allowed us to determine whether the reference frame varied as a function of stimulus category. To anticipate, our results indicate that the PPA and TOS respond to both kinds of stimuli in eye-centered rather than screen-centered coordinates, whereas RSC exhibits little evidence of visuotopic organization in either reference frame.

Methods

Subjects

Ten subjects (5 females, aged 20–28 years) were recruited from the University of Pennsylvania community and gave written informed consent in compliance with procedures approved by the University of Pennsylvania Institutional Review Board. Subjects had normal or corrected-to-normal vision. Only correction by contact lenses was permitted since these did not interfere with the eye tracking procedures explained below. An additional subject was also scanned as a pilot subject to test eye tracking but was excluded from the study prior to data analysis. Subjects were paid for their participation.

Apparatus

fMRI data were acquired at the Center for Functional Neuroimaging at the Hospital of the University of Pennsylvania on a 3-T Siemens Trio equipped with an eight-channel multiple array Nova Medical head coil. Structural T1-weighted images for anatomical localization were acquired using a 3D MPRAGE pulse sequence (TR = 1620 ms, TE = 3 ms, time to inversion = 950 ms, voxel size = 0.9766 × 0.9766 × 1 mm, matrix size = 192 × 256 × 160). T2*-weighted images sensitive to blood oxygenation level dependent (BOLD) contrasts were acquired using a gradient-echo echo-planar pulse sequence (time repetition [TR] = 3000 ms, time echo [TE] = 30 ms, voxel size = 3 × 3 × 3 mm, matrix size = 64 × 64 × 45).

Visual stimuli were rear-projected onto a Mylar screen at the head of the scanner with an Epson 8100 3-LCD projector equipped with a Buhl long-throw lens and viewed through a mirror mounted to the head coil. The entire projected field subtended 22.9 × 17.4° and was viewed at 1024 × 768 pixel resolution. Responses were recorded using a 4-button fiber-optic response pad system.

To ensure subjects maintained fixation, eye tracking was performed concurrent with the MRI scans using an ASL EYE-TRAC6 unit (Applied Science Laboratories, Bedford, MA) using bright pupil optics. Before scanning began, each subject fixated on each element in a 3 × 3 array of fixation targets in order to calibrate the control unit. Eye tracking data were collected for 6 of the 10 subjects; technical problems with the unit prevented data from being recorded for the remaining four subjects. However, gaze position was visually monitored throughout the scans for all subjects to ensure that proper fixation was maintained.

Procedure

The scanning session for each subject consisted of six experimental scans and two functional localizer scans.

Experimental scans were 6 min and 33 s long and consisted of an initial 15 s of fixation, eighty 3-s stimulus trials during which scenes or objects were shown, 40 interspersed 3-s null (fixation only) trials in which no

stimulus was shown, and 18 s of fixation at the end of the scan. Subjects fixated on a dot that remained at a constant screen position of either −6° (left of center), 0° (center), or +6° (right of center) throughout each scan. On each stimulus trial, a color photograph of either a scene (house with scenic background) or an object (e.g., a stapler, watering can, truck, etc., with no background) was presented at one of four possible screen positions relative to the screen center: 9° to the left, 3° to the left, 3° to the right, or 9° to the right (Figure 1). Each photograph subtended 5° horizontally and 6° vertically, at a resolution of 224 × 264 pixels. Photographs blinked on and off four times (500 ms on, 250 ms off) at the same screen position during each 3-s trial. For scans 1–3, subjects used a button box to report the color of the fixation dot, which could randomly change between white and black every 950 ms (or three times per trial). For scans 4–6, subjects maintained fixation but performed no explicit task. Within each scan, 10 different scenes and 10 different objects were presented at each of the four screen positions for a total of 80 stimulus trials per scan; 120 images of scenes and 120 images of objects were used across scans 1–3 and repeated in scans 4–6.

Functional localizer scans were 7 min and 48 s per run and consisted of 18-s blocks during which subjects viewed color photographs of places, faces, objects, and scrambled versions of objects and performed a 1-back task, as described previously (Epstein, Higgins, & Thompson-Schill, 2005).

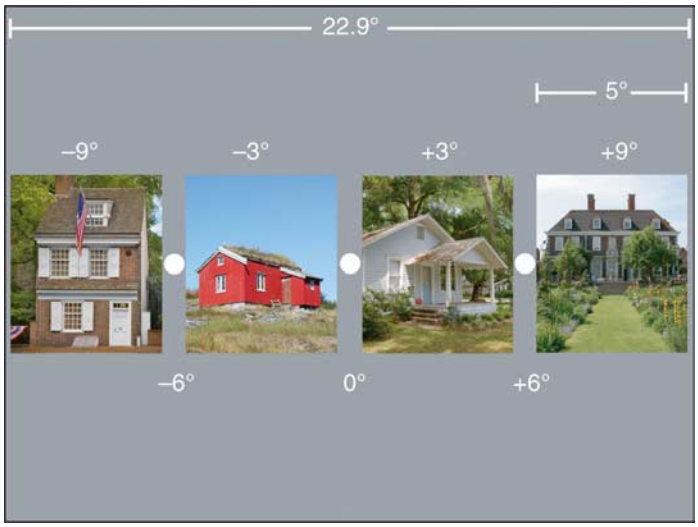


Figure 1. All possible fixation (white circles) and image positions. The entire screen subtended 22.9°, and images subtended 5° × 6°. Subjects attended to fixation dot at −6°, 0°, or +6°, while color photographs of scenes and objects were presented at one of four screen locations (−9°, −3°, +3°, +9°) per trial. Images flashed on and off four times within each 3000-ms trial. Images were scenes (shown here) and isolated objects.

Data analysis

Functional images were corrected for differences in slice timing by resampling slices in time to match the first slice of each volume, realigned with respect to the first image acquired during a scanning session, spatially normalized to the Montreal Neurological Institute (MNI) template, and then spatially smoothed with a 6-mm full-width half-maximum Gaussian filter. Data were analyzed using the general linear model as implemented in VoxBo (www.voxbo.org) including an empirically derived 1/*f* noise model, filters that removed high and low temporal frequencies, regressors to account for global signal variations, and nuisance regressors to account for between-scan differences. Data from each run were scaled to a mean of 100 before analysis to ensure that beta values were interpretable as percent signal change. Twenty-four regressors were used to model the conditions of interest (3 fixation positions × 2 stimulus types × 4 stimulus positions). Each stimulation trial was modeled as a 3-s-long event that was convolved with a standard hemodynamic response function and assigned to one of the 24 regressors. "Null" trials were not modeled and thus provided a fixation-only baseline. Both region of interest (ROI) and whole-brain analyses were performed.

For ROI analyses, scene- and object-responsive areas were identified for each subject using data from the functional localizer scans. Scene-responsive ROIs included the parahippocampal place area (PPA), retrosplenial complex (RSC), and transverse occipital sulcus (TOS), which were defined as sets of contiguous voxels with significantly greater responses to scenes than to common objects. For purposes of comparison, object- and face-responsive ROIs were also identified. These were the lateral occipital complex (LOC) defined as the set of contiguous voxels in the lateral/ventral occipitotemporal region that responded more to objects than to scrambled objects and the fusiform face area (FFA) defined as the set of contiguous voxels in the middle fusiform gyrus that responded more to faces than to objects. The significance threshold defining these ROIs was set on a subject-by-subject basis so that ROIs were consistent with those identified in previous studies (Epstein et al., 1999; Epstein & Kanwisher, 1999; Epstein, Parker, & Feiler, 2007); thresholds were either $t > 3.0$ or $t > 3.5$. Using these criteria, we were able to define ROIs in the following number of subjects: PPA (10 L, 10 R), RSC (10 L, 10 R), TOS (10 L, 10 R), LOC (10 L, 10 R), FFA (7 L, 10 R). The time course of MR response during the experimental scans was then extracted from each ROI (averaging over all voxels) and entered into a general linear model used to calculate parameter estimates (beta values) for the 24 conditions of interest. Extracted beta values were then used as dependent variables in a second-level random effects analysis of variance.

For whole-brain analyses, subject-specific beta maps were calculated for contrasts of interest and then

smoothed to 10-mm FWHM to facilitate between-subject averaging before entry into a random effects analysis. Voxels with a significance level of $P < 0.01$, uncorrected, are reported.

Results

Our primary objective was to determine whether the spatial topography of scene-selective regions (PPA, TOS, RSC) is expressed in eye-centered or screen-centered coordinates. Visually evoked responses in these regions for centrally fixating participants tend to be higher for stimuli appearing in the contralateral visual field (MacEvoy & Epstein, 2007). The critical question is whether this contralateral bias reflects a preference for stimuli contralateral to the fixation point (eye-centered coordinates) or contralateral to the center of the screen (screen-centered coordinates). Note that because the head and body (trunk) positions are fixed within the scanner bore, screen-centered coordinates are indistinguishable from head- or trunk-centered coordinates in the current experiment.

For each ROI, we computed fMRI response as a function of stimulus position expressed relative to either the center of gaze or to the center of the stimulus screen (see Figure 2). Separate response curves were plotted for each of the three fixation positions in each of these reference frames. Percent signal change was computed by comparing response on experimental trials to null/fixation trials. If a region encodes information in eye-centered coordinates, then the three curves should align in the eye-centered reference frame but be offset in the screen-centered frame. Alternatively, if a region encodes information in screen-centered coordinates, then the three curves should align in the screen-centered frame but be offset in the eye-centered frame.

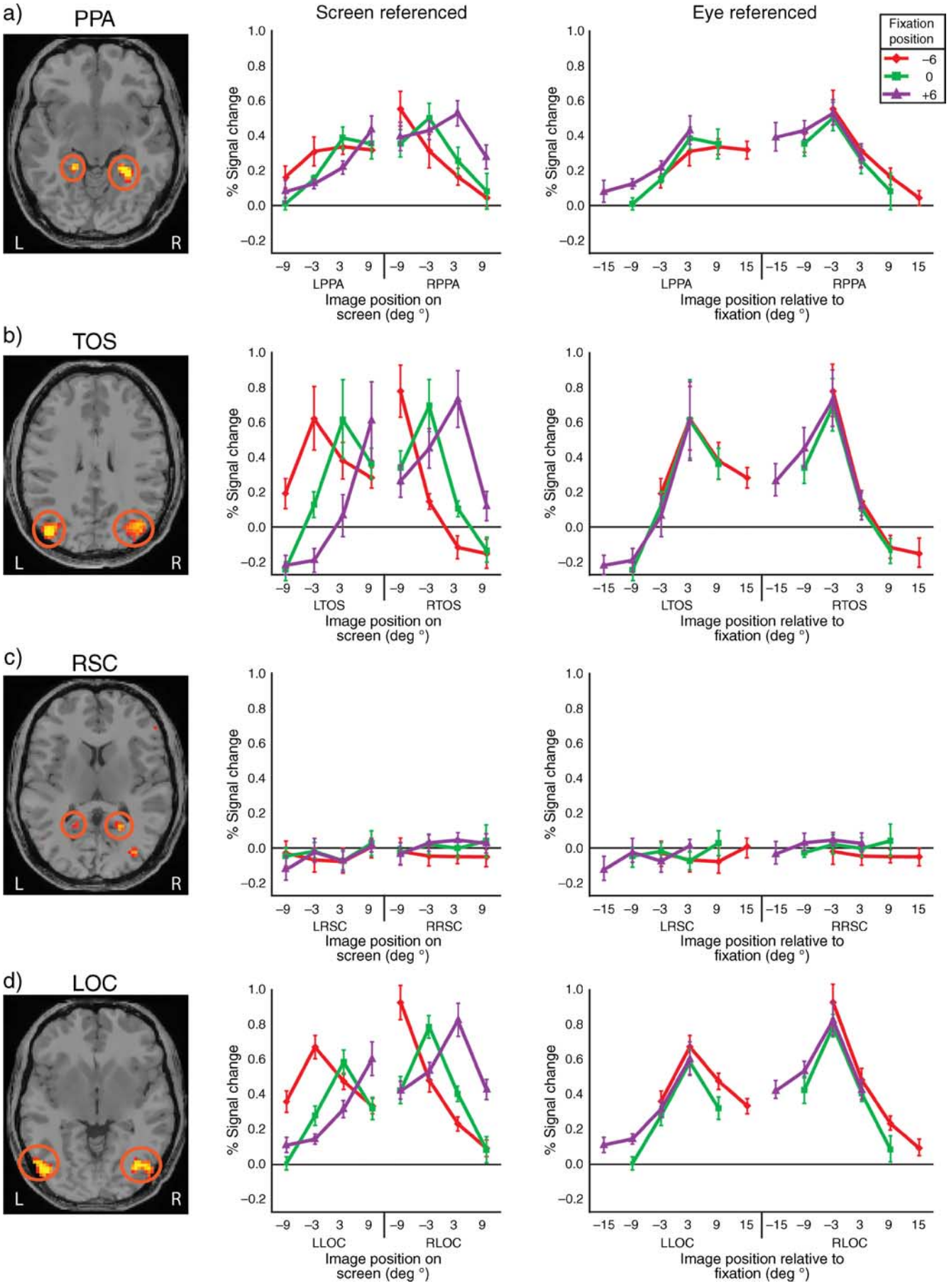
To assess the degree of alignment between the three response curves in each reference frame, we performed separate repeated-measures analyses of variance (ANOVAs) for each reference frame with stimulus type, stimulus position, and fixation position as factors. In this analysis, a meaningful offset of the response curves between fixation positions is indicated by a significant stimulus position × fixation position interaction. Because only the middle two image positions (−3° and 3°) occur at all three fixation positions when the activation curves are shifted into eye-centered coordinates, we restricted our analyses to the middle two stimulus positions in both cases, thus ensuring a fair comparison between eye-centered and screen-centered reference frames. Below we report the results of these analyses for the three scene-selective ROIs (PPA, TOS, RSC) one object-selective ROI (LOC) and one face-selective ROI (FFA) along with a related contrast performed across the entire brain.

Parahippocampal place area

Consistent with previous results (Arcaro et al., 2009; MacEvoy & Epstein, 2007), PPA response was significantly higher for contralateral compared to ipsilateral stimuli when fixation was at screen center (left hemisphere: $t(9) = 4.45$, $P = 0.0016$; right hemisphere: $t(9) = 7.55$, $P = 0$ 0.000035). Critically, this contralateral bias is expressed in eye-centered coordinates: when response curves are plotted in screen-centered coordinates, they are clearly offset (Figure 2a, left panel). In contrast, the curves fall into close alignment when stimulus position is plotted in eye-centered coordinates (Figure 2a, right panel). The results of the ANOVA supported these observations. There was a significant interaction of stimulus position and fixation position when stimulus position was defined in screen-centered coordinates (left: $F(2,18) = 5.072$, $P = 0.018$; right: $F(2,18) = 16.652$, $P = 0.0001$), but no interaction when stimulus position was defined in eye-centered coordinates (left: $F(1,18) = 0.684$, $P = 0.517$; right: $F(2,18) = 0.170$, $P = 0.845$). In other words, the PPA responds more strongly to stimuli presented contralateral to fixation and less strongly to stimuli presented ipsilateral to fixation, regardless of the location of the fixation point.

These results were found to be independent of task, which differed between scans 1–3 (report color of the fixation dot) and scans 4–6 (passive viewing). Recently, Crespi et al. (2009) and Morrone, Cicchini, and Burr (2010) have suggested that reference frames may depend on attentional differences between passive fixation and performing a fixation task. To determine whether task differences affected our results, we passed our fMRI data through a new general linear model in which beta values for each condition were ascertained separately for the two halves of the experiment. An ANOVA was then run on these beta values with stimulus position, fixation position, and task as factors in both eye-centered and screen-centered coordinates. We found no significant stimulus position × fixation position × task interaction in either

Figure 2. fMRI responses in (a) PPA, (b) TOS, (c) RSC, and (d) LOC to images of scenes and objects presented at four screen locations while subjects fixated at one of three different screen locations relative to center (−6°, 0°, +6°); error bars are ±1 standard error of the mean for each condition. On the left, response curves plotted as a function of subtended distance from the center of the screen (taken as 0°) to the center of the image (screen-centered coordinates). On the right, the same response curves plotted as function of subtended distance from fixation (taken as 0°) to the center of image (eye-centered coordinates). The response curves in the PPA, TOS, and LOC were offset when plotted in screen-centered coordinates but overlapped when plotted in eye-centered coordinates, a pattern indicative of retinotopic (eye-centered) encoding.


a)
PPA
b)
TOS
c)
RSC
d)
LOC
L
R
Screen referenced
Eye referenced
Fixation position
−6
0
+6
% Signal change
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−9 −3 3 9 −9 −3 3 9
−15 −9 −3 3 9 15 −15 −9 −3 3 9 15
LPPA RPPA
LTOS RTOS
LRSC RRSC
LLOC RLOC
Image position on screen (deg °)
Image position relative to fixation (deg °)

reference frame (eye-centered: $F(2,14) = 0.722$, $P = 0.503$, screen-centered PPA: $F(2,14) = 2.155$, $P = 0.153$).

The above data reflect the average of the scene and object responses, but we observed similar patterns when we considered the response to scenes alone. Specifically, there was a significant stimulus position × fixation position interaction when scene position was defined in the screen-centered reference frame (left: $F(2,18) = 6.389$, $P = 0.008$; right: $F(2,18) = 8.574$; $P = 0.002$) but not when it was defined in the eye-centered reference frame (left: $F(2,18) = 2.203$, $P = 0.139$; right: $F(2,18) = 0.575$, $P = 0.573$; Figure 3a). For objects, the stimulus position × fixation position interaction in the screen-centered reference frame was only marginally significant in the right hemisphere ($F(2,18) = 3.314$, $P = 0.060$) and non-significant in the left hemisphere ($F(2,18) = 0.390$, $P = 0.683$). Although this last result might suggest that scenes and objects are encoded under different reference frames, it is also possible that responses to objects were simply noisier. Indeed, we observed no significant fixation position × stimulus position interaction for objects when stimulus location was defined in eye-centered coordinates (left: $F(2,18) = 0.398$, $P = 0.677$; right: $F(2,18) = 0.113$, $P = 0.894$; Figure 3a). The most parsimonious conclusion is that the PPA encodes both scenes and objects in eye-centered coordinates.

Transverse occipital sulcus

Patterns of activation in TOS were similar to those observed in the PPA. When gaze was at screen center, TOS responses showed a significant contralateral bias (left: $t(9) = 3.77$, $P = 0.00438$; right: $t(9) = 4.75$, $P = 0.00104$). As in the PPA, this bias appeared to reflect eye-centered rather than scene-centered coding. There was a significant interaction of stimulus position and fixation position when position was defined in screen-centered coordinates (left: $F(2,18) = 6.130$, $P = 0.009$; right: $F(2,18) = 13.994$, $P = 0.000216$) but not when position was defined in eye-centered coordinates (left: $F(2,18) = 0.142$, $P = 0.868$; right: $F(2,18) = 0.066$, $P = 0.936$). This difference is illustrated clearly in Figure 2b; position–response curves for each fixation position are shifted for stimuli in screen-centered coordinates (left) but overlap almost perfectly in eye-centered coordinates (right). As in the

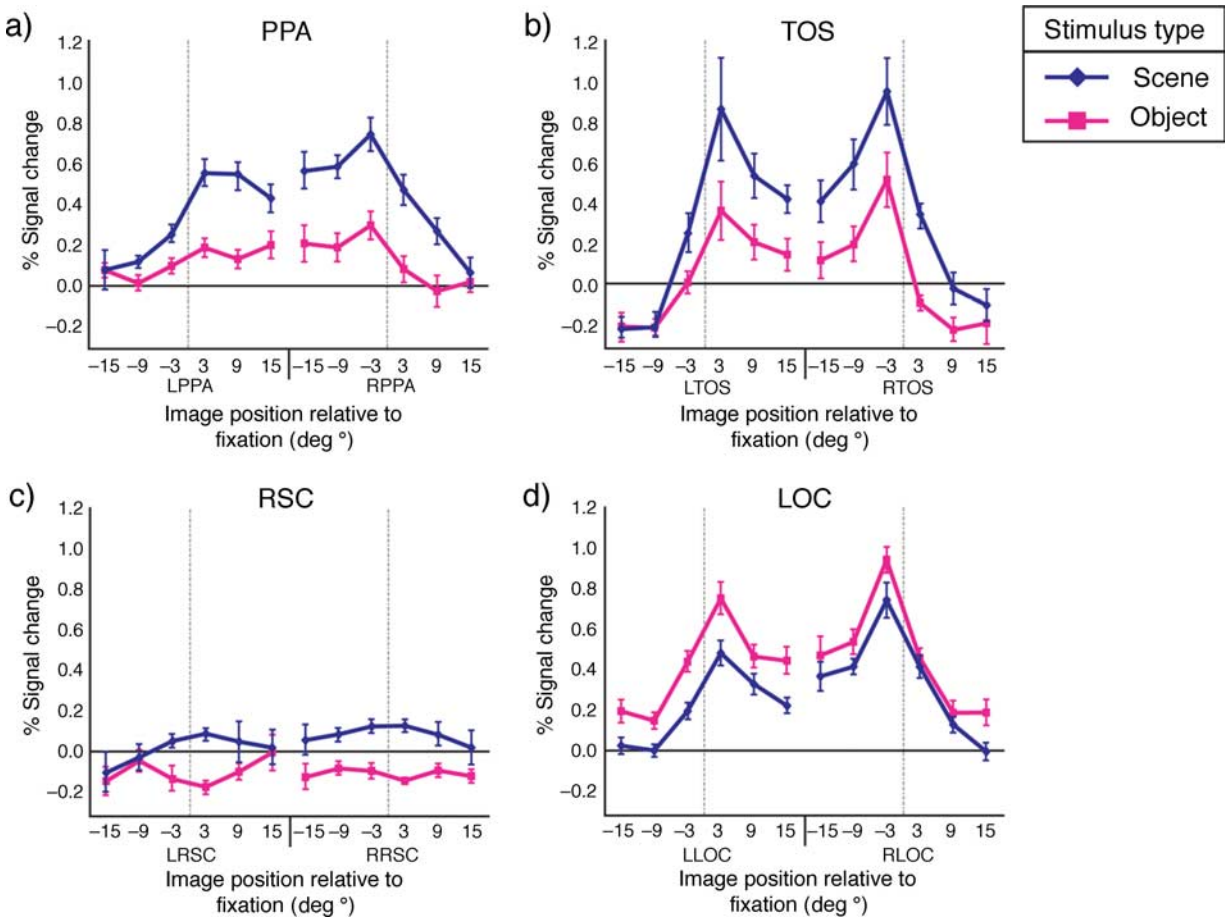


Figure 3. fMRI response in (a) PPA, (b) TOS, (c) RSC, and (d) LOC to images of scenes and objects presented at six different positions relative to fixation. To a first approximation, PPA, TOS, and LOC appear to encode both scenes and objects in the same reference frame. Hatched gray lines indicate fixation position. Error bars are ±1 standard error of the mean for each condition.

PPA, these results were found to be independent of task (Fs < 2, Ps > 0.18).

We observed similar results when TOS responses to scenes and objects were considered separately. For scenes, the fixation position × stimulus position interaction was significant in screen-centered coordinates (left: $F(2,18) = 7.153$, $P = 0.005$; right: $F(2,18) = 17.391$, $P = 0.000062$) but not in eye-centered coordinates (left: $F(2,18) = 0.057$, $P = 0.944$; right: $F(2,18) = 0.738$, $P = 0.492$; Figure 3b). For objects, the results were similar but less reliable: the fixation position × stimulus position interaction was significant in screen-centered coordinates for the right hemisphere ($F(2,18) = 7.917$, $P = 0.003$) but fell short of significance in the left hemisphere ($F(2,18) = 2.659$, $P = 0.097$); the equivalent interaction was non-significant in both hemispheres for eye-centered coordinates (left: $F(2,18) = 0.329$, $P = 0.724$; right: $F(2,18) = 0.230$, $P = 0.797$; Figure 3b). In sum, our results clearly indicate eye-centered encoding for scenes and suggest that objects are probably encoded using the same reference frame.

Retrosplenial complex

Figure 2c shows responses in RSC as a function of stimulus position expressed in screen-centered (left) and eye-centered (right) coordinates. In contrast to the PPA and TOS, RSC did not show response that varied as a function of stimulus position when fixation was at screen center (left: $t(9) = 0.333$, $P = 0.747$; right: $t(9) = 0.025$, $P = 0.981$). Indeed, RSC did not respond significantly above baseline for any stimulus position. In the absence of any contralateral bias, the question of reference frame is unanswerable. We obtained the same result when we considered responses to objects and scenes separately (Figure 3c).

Object- and face-responsive regions

For purposes of comparison, we also examined visuotopic responses in the lateral occipital complex (LOC), which was functionally defined by significantly higher responses to images of objects than to images of scrambled objects (Malach et al., 1995), and in the fusiform face area (FFA), which was functionally defined by higher response to images of faces than of objects (Kanwisher, McDermott, & Chun, 1997).

Previous studies have indicated that response curves in LOC are retinotopic (eye-centered) rather than spatiotopic (screen-centered; Gardner et al., 2008), and our results were consistent with these earlier findings. LOC responded significantly more strongly to contralateral than to ipsilateral stimuli when fixation was at screen center (left: $t(9) = 7.44$, $P = 0.0000394$; right: $t(9) = 8.40$, $P = 0.0000150$). As in the PPA and TOS, this contralateral bias appears to reflect eye-centered encoding, as position–response curves for each eye position were offset when positions were defined in screen-centered coordinates but overlapped closely when eye-centered coordinates were used (Figure 2d). There was a significant stimulus position × fixation position interaction for screen-centered coordinates (left: $F(2,18) = 28.026$, $P = 0.000003$; right: $F(2,18) = 28.437$, $P = 0.000003$) but no interaction for eye-centered coordinates (left: $F(2,18) = 0.283$, $P = 0.757$; right: $F(2,18) = 1.097$, $P = 0.355$). These results are clearly consistent with an eye-centered reference frame. As in the PPA and TOS, these results were independent of task (Fs < 1.2, Ps > 0.3).

The same eye-centered reference frame appeared to be used for both scenes and objects (Figure 3d). In screen-centered coordinates, there was a significant interaction of stimulus position and fixation position (indicating response curve offset) for both objects (left: $F(2,18) = 22.594$, $P = 0.000012$; right: 19.841, $P = 0.000028$) and scenes (left: $F(2,18) = 6.709$, $P = 0.007$; right: $F(2,18) = 9.026$, $P = 0.002$). In eye-centered coordinates, on the other hand, there was no significant stimulus position × fixation position interaction for either objects (left: $F(2,18) = 0.017$, $P = 0.983$; right: $F(2,18) = 3.346$, $P = 0.058$) or scenes (left: $F(2,18) = 0.931$, $P = 0.412$; right: $F(2,18) = 1.603$, $P = 0.229$).

LOC can be divided into two subdivisions based on the tendency of object-selective activity to localize at two foci along the rostrocaudal axis (Grill-Spector, 2003; MacEvoy & Epstein, 2007; Malach, Levy, & Hasson, 2002). The more anterior/medial subdivision is associated with the posterior fusiform gyrus (pF) while the posterior/lateral subdivision is associated with the latter occipital cortex per se (LO). Separate analysis performed on these two subdivisions revealed eye-centered coding in both.

In the FFA, interaction of stimulus position and fixation position when stimulus position was defined in screen-centered coordinates was significant for the right hemisphere ($F(2,14) = 14.060$, $P = 0.0004$) and marginally significant for the left hemisphere ($F(2,10) = 3.506$, $P = 0.070$). In contrast, there was no interaction in either hemisphere when stimulus position was defined in eye-centered coordinates (left: $F(2,10) = 0.868$, $P = 0.449$; right: $F(2,18) = 0.013$, $P = 0.987$). Similar results were found when responses to scenes and responses to objects were analyzed separately. Thus, similar to the PPA, TOS, and LOC, FFA exhibits an eye-centered response profile, at least for the non-preferred stimuli (scenes and objects) examined here.

Whole-brain analysis

We performed additional analyses to identify areas outside our functionally defined ROIs that exhibited responses in screen-centered or eye-centered coordinates. Screen-centered and eye-centered voxels were defined by

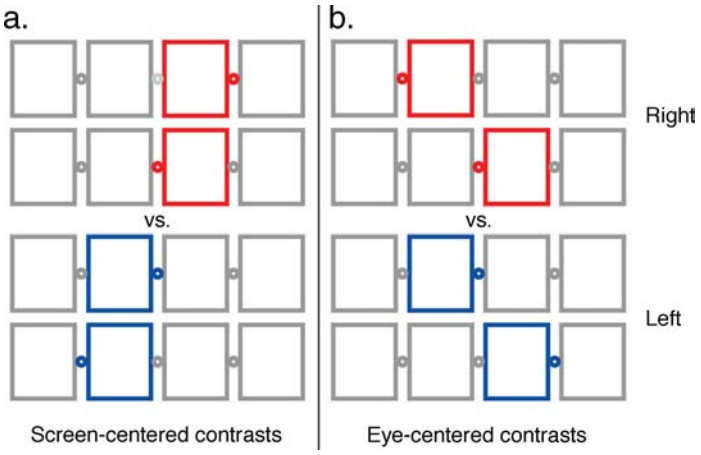


Figure 4. Contrasts used in the whole-brain analysis. Stimulus positions are indicated by the colored boxes; fixation positions are indicated by the colored annuli. (a) To find regions with a left–right difference in screen-centered coordinates, we grouped trials to introduce an overall difference in screen position with no overall difference in eye-centered position. Specifically, response to stimuli presented at screen position +3 (red) was compared to response to stimuli presented at screen position −3 (blue), averaging over trials in which stimuli appeared either to the left or right of fixation. (b) To find regions with a left–right difference in eye-centered coordinates, we grouped trials to introduce an overall difference in fixation-referenced position with no overall difference in screen position. Specifically, response to stimuli presented at screen positions +3 and −3 with fixation on the left (red) was compared to response to stimuli presented at the same screen positions with fixation on the right.

examining differential response for left vs. right presentations, where left and right were defined in one reference frame while controlling for left vs. right differences in the other frame. The logic behind this analysis is illustrated in Figure 4. The left side of the figure illustrates the contrast used to define screen-centered voxels, whereas the right side of the figure illustrates the contrast used to define eye-centered voxels. For the screen-centered contrast, activity evoked by stimuli presented at screen position +3 was contrasted with activity evoked by stimuli presented at screen position −3; however, this contrast was restricted to trials in which the fixation position was either directly to the left or directly to the right of the stimulus in order to balance stimulus positions in eye-centered coordinates (Figure 4a). For the eye-centered contrast, stimuli presented at screen positions −3 and +3 with fixation directly to the left were contrasted with stimuli presented at −3 and +3 with fixation directly to the right. Here, screen position was balanced, while there was an overall difference in eye-centered position (Figure 4b).

Despite the use of a relatively liberal significance threshold, we observed no regions exhibiting response curves defined in a screen-centered reference frame. In contrast, large swaths of occipitotemporal cortex showed evidence of eye-centered encoding (Figure 5). Whole-brain scene-selective ROIs (PPA, RSC, TOS) were identified based on anatomical overlap of the individual subject ROIs; voxels shared across half or more subjects were included. Regions showing eye-centered encoding overlapped with TOS and the posterior portion of the PPA. The general picture is of two eye-centered processing streams that originate in V1 and have the TOS and PPA as endpoints.

Eye tracking data

To confirm the reliability of subject fixation, we constructed histograms of fixation positions for the 6/10 subjects for which these data were recorded. For each subject, we identified fixation periods as all non-overlapping 50-ms intervals within which gaze deviated no more than 0.5 degree from its position at the beginning of the interval. Intervals with starting positions >20 degrees

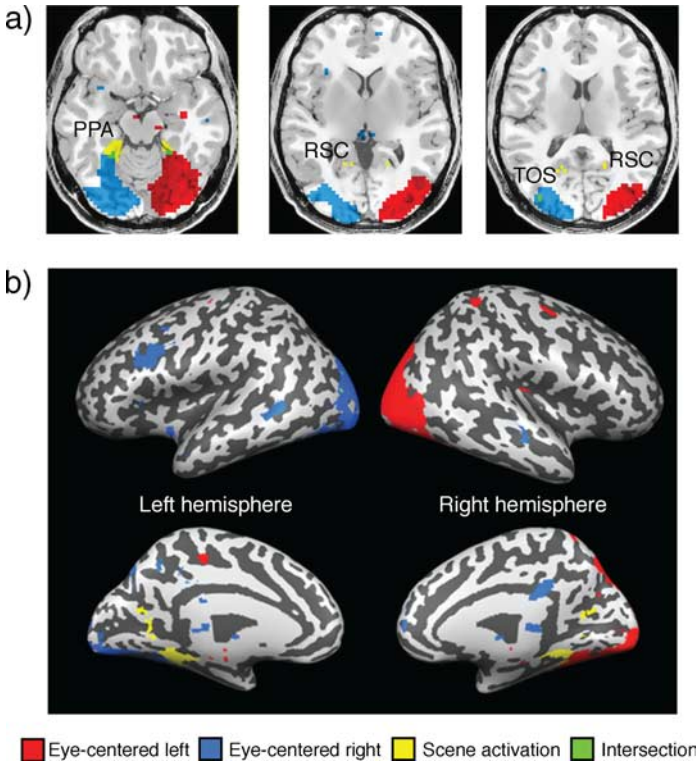


Figure 5. Whole-brain analysis. Regions showing a left (red)–right (blue) difference in eye-centered coordinates are plotted with regions responding more to scenes than objects (yellow). The intersection of these data is shown in green. (a) Posterior PPA and TOS overlap with the eye-centered regions, but RSC (middle and right) does not. Scene-selective regions (PPA, RSC, TOS) were identified based on anatomical overlap of the individual subject ROIs; voxels shared across half or more subjects were included. (b) Inflated brain illustrating eye-centered regions relative to scene-selective regions.

from screen center were discarded as these reflect loss of signal rather than true eye position. Supplementary Figure 1 shows the fixation position histogram for a typical subject, plotted in screen coordinates. Fixation positions cluster around the fixation points (−6°, 0°, and 6°) with no peaks at the positions of the visual stimuli (−9°, −3°, 3°, and 9°). To quantify fixation performance, we measured the absolute value of the difference between each fixation interval's starting point and the position of the fixation target at that moment. When accumulated across all subjects, the mean deviation from target was 0.91° with a standard deviation of 0.91°. When fixation deviations were averaged within subjects, the mean deviation across subjects was 1.24°, with a standard deviation of 1.08°.

Discussion

In this study, we determined the reference frame of spatial topography in scene-selective regions by analyzing responses to stimuli presented at different screen positions while gaze position was varied. A region encoding stimulus position in a screen-centered frame will respond equally to stimuli presented at the same screen location, regardless of where the eyes are oriented. In the fMRI apparatus, the relative positions of the screen, head, and body are fixed, so a screen-centered response could reflect head-, trunk-, or world-centered coding. An eye-centered region, on the other hand, will respond equally to stimuli only when they are presented at the same position relative to the position of the eyes (fixation). We have confirmed that object-selective LOC exhibit eye-centered response curves (McKyton & Zohary, 2007) and we have determined that the PPA and TOS also exhibit eye-centered response curves. In the retrosplenial cortex (RSC), there was little discernible variation in response across stimulus position, making the question of reference frame moot (at least for the range of stimulus positions we used). Whole-brain analysis revealed no regions exhibiting screen-centered response but eye-centered responses in a large region extending from the early visual cortex to the TOS and posterior PPA.

Previous fMRI studies have examined the effect of presenting images at different visual field locations on PPA, RSC, and TOS responses. For example, in an earlier study, we presented scenes and objects on either the left or right side of the fixation point (MacEvoy & Epstein, 2007). fMRI response in the PPA and TOS was greater for contralateral compared to ipsilateral presentation (e.g., greater response in the left PPA to scenes appearing in the right visual field), although the response to ipsilateral presentation was still considerably above baseline. Consistent with the large ipsilateral response, cross-position fMRI adaptation (i.e., fMRI response attenuation caused by repeating stimuli in different hemifields) was just as strong as same-position adaptation (i.e., fMRI response reduction caused by repeating stimuli in the same hemifield), suggesting that the PPA and TOS support receptive fields large enough to extend across the vertical meridian. These findings were supported and elaborated by the results of a recent study by Arcaro et al. (2009) that used standard retinotopic mapping paradigms to identify two new visual maps (PH-1 and PH-2) in the parahippocampal region. These authors speculated that these maps were missed by earlier studies because they contain neurons with large receptive fields that are only weakly modulated by stimulus position. Interestingly, little evidence for topographical organization has been observed in RSC. Indeed, in our 2007 study, RSC showed little response of any sort to off-centered presentation, a finding that we replicate here (Figure 3c).

Importantly, earlier studies did not examine the reference frame used by these areas to encode visual information. Determining the reference frame of visually evoked activity differs from visuotopic mapping. Although several mapping studies using fMRI have shown the existence of an orderly representation of visual space in a variety of human extrastriate areas (Engel et al., 1994; Sereno et al., 1995; Wandell et al., 2007), most have used a paradigm in which the subject is required to maintain central fixation. This leaves open the question of whether these visuotopic maps are yoked to fixation position and are updated with each eye movement, or whether they are fixed to more stable references, such as the body, head, or external objects. Differentiating among these possibilities requires varying gaze position in order to dissociate eye-centered coordinates from the others, as we have done here.

The selectivity of PPA for scenes sets it apart from other visual areas and theoretically makes it a good candidate to host a non-eye-centered reference frame. In contrast to human interactions with objects—which typically only involve objects within the visual field—navigation takes place within an extended environment that exceeds what the visual system captures at any given eye position. Therefore, regions putatively involved in navigation would seem to benefit from visual representations of visual space that are stable across eye movements. Despite this logic, our results demonstrate the persistence of an eye-centered reference frame in the PPA, consistent with recent proposals that the PPA is important for "online" analysis of the visual features of a scene (Epstein, 2008). Furthermore, we found no evidence for a spatiotopic response in the PPA tied to the head, body, or screen. Thus, our results extend earlier findings indicating eye-centered encoding across the ventral visual stream (Gardner et al., 2008, but see d'Avossa et al., 2007).

Do our results imply that *all* neurons in the PPA exhibit eye-centered RFs? Not necessarily. Indeed, we can think of three possible alternative scenarios. First, eye-centered responses in the PPA might be modulated by head and

body positions using a gain field mechanism similar to that observed in parietal cortex (Andersen, Essick, & Siegel, 1985; Snyder, Grieve, Brotchie, & Andersen, 1998). We were not able to investigate this possibility in the current experiment because head, body, and screen positions could not be manipulated relative to each other. Second, more spatially precise analyses of response patterns might reveal evidence that some PPA neurons have eye-centered RFs while others have head- or body-centered RFs. This possibility could be probed through fMRI repetition suppression, in which the response of distinct neural populations can be attenuated or suppressed by repeating stimuli under different conditions (Grill-Spector & Malach, 2001). For example, if the PPA contains neurons with both eye-centered and head-centered RFs, then repetition of stimuli at the same screen position (while varying fixation) may yield stronger signal suppression than repetition of visual features at different screen positions. Finally, some neurons in the PPA might encode spatial quantities that are intrinsic to the visual scene and thus "move" when the stimulus is shifted along the screen. The presence of such scene-centered neurons could account for the position-invariant repetition suppression observed in the PPA when scenes were repeated at different screen locations in an earlier study (MacEvoy & Epstein, 2007) as well as reports that the PPA is especially engaged when subjects make judgments about the positions of objects in scene-centered coordinates (Committeri et al., 2004; Galati, Pelle, Berthoz, & Committeri, 2010). In this scenario, the PPA may play a role in both visual recognition of scenes and extraction of spatial information from scenes, the former involving neurons with eye-centered RFs and the latter involving transfer of information between eye-centered and scene-centered neurons. The integration of eye-centered and scene-centered information in the PPA may also contribute to a robust representation of perceived position vs. actual location, demonstrated recently by Fischer, Spotswood, and Whitney (2011).

In contrast to the PPA (and TOS), we observed no evidence for eye-centered coding, or position-dependent responses of any sort, in RSC. Although this finding may be a result of the relatively circumscribed set of stimulus positions we used, which were small relative to the full expanse of the visual field, it is also consistent with proposals that RSC is central for the task of situating the currently viewed scene within the broader spatial environment (Epstein, 2008). Eye-centered encoding in the RSC would be undesirable in this view because the information that it encodes (the position and orientation of the observer relative to a spatial frame that extends beyond the local scene) does not change with eye movements. Indeed, consistent with earlier studies (MacEvoy & Epstein, 2007), we observed little response of any kind in RSC to stimuli presented offset from fixation, suggesting that this region only engages when stimuli are centrally presented and thus more likely to be treated not just as a scene but as a "place." Taken as a whole, these findings highlight the putative functional distinction between the PPA and RSC (Epstein et al., 2007), whereas the PPA appears to play a critical role in scene perception, RSC may support retrieval of spatial representations from memory, a function that does not require neurons with visuotopic response.

In sum, our data indicate eye-centered encoding of visual information in the PPA, TOS, and LOC but not RSC. These results suggest that eye-centered encoding is an important common principle that is obtained in both scene- and object-responsive areas of the visual system.

Acknowledgments

This work was supported by grants from the National Science Foundation (SBE-0541957) and National Institutes of Health (EY-016464) to R.A.E. The authors wish to thank Mary Smith and three anonymous reviewers for their suggestions.

Commercial relationships: none.
Corresponding author: Emily J. Ward.
Email: emily.ward@yale.edu.
Address: 3720 Walnut St. Room B51, Philadelphia, PA 19104, USA.

References

Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: A synthesis and taxonomy. *Brain, 122,* 1613–1628.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron, 21,* 373–383.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230,* 456–458.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience, 29,* 10638.

Committeri, G., Galati, G., Paradis, A. L., Pizzamiglio, L., Berthoz, A., & LeBihan, D. (2004). Reference frames for spatial cognition: Different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location. *Journal of Cognitive Neuroscience, 16,* 1517–1535.

Crespi, S., Biagi, L., Burr, D. C., d'Avossa, G., Tosetti, M., & Morrone, M. C. (2009). Spatial attention modulates the spatiotopicity of human MT complex. *Perception, 38,* ECVP Abstract Supplement.

d'Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience, 10,* 249–255.

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E., et al. (1994). fMRI of human visual cortex. *Nature, 369,* 525–525.

Epstein, R. A. (2005). The cortical basis of visual scene processing. *Visual Cognition, 12,* 954–978.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences, 12,* 388–396.

Epstein, R. A., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area. *Neuron, 23,* 115–125.

Epstein, R. A., Higgins, J. S., & Thompson-Schill, S. L. (2005). Learning places from views: Variation in scene processing as a function of experience and navigational ability. *Journal of Cognitive Neuroscience, 17,* 73–83.

Epstein, R. A., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Epstein, R. A., & Kanwisher, N. (1999). Repetition blindness for locations: Evidence for automatic spatial coding in an RSVP task. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1855–1866.

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *Journal of Neuroscience, 27,* 6141.

Fischer, J., Spotswood, N., & Whitney, D. (2011). The emergence of perceived position in the visual system. *Journal of Cognitive Neuroscience, 23,* 119–136.

Galati, G., Pelle, G., Berthoz, A., & Committeri, G. (2010). Multiple reference frames used by the human brain for spatial perception and memory. *Experimental Brain Research, 206,* 109–120.

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience, 28,* 3988.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology, 13,* 159–166.

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

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica, 107,* 293–321.

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

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 9379–9384.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293,* 1506–1509.

MacEvoy, S. P., & Epstein, R. A. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *Journal of Neurophysiology, 98,* 2089.

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *The Journal of Psychology, 42,* 225–238.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., & O'Keefe, J. (1998). Knowing where and getting there: A human navigation network. *Science, 280,* 921–924.

Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences, 6,* 176–184.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America, 92,* 8135–8139.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences, 200,* 269–294.

McKyton, A., & Zohary, E. (2007). Beyond retinotopic mapping: The spatial representation of objects in the human lateral occipital complex. *Cerebral Cortex, 17,* 1164.

Mendez, M. F., & Cherrier, M. M. (2003). Agnosia for scenes in topographagnosia. *Neuropsychologia, 41,* 1387–1395.

Morrone, M. C., Cicchini, M., & Burr, D. C. (2010). Spatial maps for time and motion. *Experimental Brain Research, 206,* 121–128.

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., et al. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing: A PET study. *Brain, 123,* 1903–1912.

Sayres, R., & Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. *Journal of Neurophysiology, 100,* 249–267.

Sereno, M., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268,* 889–893.

Snyder, L. H., Grieve, K. L., Brotchie, P., & Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature, 394,* 887–891.

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