

Full Scenes produce more activation than Close-up Scenes and Scene-Diagnostic Objects in parahippocampal and retrosplenial cortex: An fMRI study

John M. Henderson ^{a,*}, Christine L. Larson ^b, David C. Zhu ^b

^a Visual Cognition Research Unit, Human Cognitive Neuroscience, Psychology, 7 George Square, University of Edinburgh, Edinburgh EH8 9JZ, UK

^b Department of Psychology, Michigan State University, East Lansing, MI 48824-1117, USA

Accepted 7 May 2007

Available online 2 July 2007

Abstract

We used fMRI to directly compare activation in two cortical regions previously identified as relevant to real-world scene processing: retrosplenial cortex and a region of posterior parahippocampal cortex functionally defined as the parahippocampal place area (PPA). We compared activation in these regions to full views of scenes from a global perspective, close-up views of sub-regions from the same scene category, and single objects highly diagnostic of that scene category. Faces were included as a control condition. Activation in parahippocampal place area was greatest for full scene views that explicitly included the 3D spatial structure of the environment, with progressively less activation for close-up views of local scene regions containing diagnostic objects but less explicitly depicting 3D scene geometry, followed by single scene-diagnostic objects. Faces did not activate parahippocampal place area. In contrast, activation in retrosplenial cortex was greatest for full scene views, and did not differ among close-up views, diagnostic objects, and faces. The results showed that parahippocampal place area responds in a graded fashion as images become more completely scene-like and include more explicit 3D structure, whereas retrosplenial cortex responds in a step-wise manner to the presence of a complete scene. These results suggest scene processing areas are particularly sensitive to the 3D geometric structure that distinguishes scenes from other types of complex and meaningful visual stimuli.

© 2007 Elsevier Inc. All rights reserved.

Keywords: fMRI; Scene processing; Real-world scenes; Object processing; Parahippocampal place area; Parahippocampal cortex; Retrosplenial cortex

1. Introduction

A real-world scene can be defined as a specific human-scaled view of the 3D environment within which we are embedded (Henderson & Ferreira, 2004; Henderson & Hollingworth, 1999). Real-world scenes differ from other types of visual stimuli in a number of important ways. For example, scenes are recognized more quickly than would be expected on the basis of an analytical route involving serial recognition of individual objects and their spatial relationships (Intraub, 1981; Potter, 1976; Schyns & Oliva, 1994; see also, Li, Van Rullen, Koch,

& Perona, 2002; Thorpe, Fize, & Marlot, 1996). Scenes possess particular kinds of spatial and semantic relationships and constraints that define the nature of the contextual and associative relationships that are possible and likely within them (Biederman, Mezzanotte, & Rabinowitz, 1982; Henderson & Hollingworth, 1999; Torralba, Oliva, Castelhano, & Henderson, 2006). These characteristics of naturalistic scenes map onto two general hypotheses that have been advanced concerning the nature of the computations and representations supported by cortical regions underlying scene processing: computation of geometric spatial structure (including the type of spatial information relevant for spatial navigation), and processing of semantic relationships and contextual associations.

* Corresponding author. Fax: +44 131 650 3461.

E-mail address: john.m.henderson@ed.ac.uk (J.M. Henderson).

Two cortical regions have recently been implicated in analyzing scene-specific information. First, the posterior region of the parahippocampal cortex (pPHC) in the medial temporal lobe has been identified with scene processing (Aguirre & D'Esposito, 1997; Bar & Aminoff, 2003; Kohler, Crane, & Milner, 2002; Maguire, Frith, & Cipolotti, 2001). More precisely, a functionally-defined region of pPHC, the parahippocampal place area (PPA), has been shown to be selectively activated by scenes over other types of visual stimuli including faces, single objects, and arrays of multiple objects (Epstein & Kanwisher, 1998). Some evidence suggests that this region responds preferentially to images that depict the 3D spatial structure of scenes over those that do not, leading to the proposal that PPA is specifically involved in processing the 3D geometry and spatial structure of local space (Epstein, 2005; Epstein, Graham, & Downing, 2003; see Epstein, 2005, for review). These results suggest that an important way in which scenes differ from other types of complex visual stimuli is by virtue of their spatial geometry.

If the 3D geometry hypothesis for PPA function is correct, we might expect that a more complete view of the local environment explicitly depicting the global 3D geometry of local navigable space would evoke a stronger PPA response than a close-up scene view that does not explicitly depict this geometry. To date, contrasts of this type have only been reported in two fMRI studies. Epstein et al. (2003) described a pilot experiment comparing close-up tabletop scenes to landscapes and rooms and reported equivalent PPA activation across them. However, the complete details of this pilot study were not reported, making it difficult to assess the conclusions. Furthermore, indoor scenes have been found to activate PPA more than outdoor scenes (Henderson, Larson, & Zhu, 2007), so the tabletop versus landscape comparison may have been susceptible to an indoor–outdoor confound. In a second report comparing close-up and full scene views, Bar and Aminoff (2003) included in two experiments images of close-up scenes containing objects with strong contextual associations to particular scenes, and indoor and outdoor control scenes. Though they did not specifically compare these conditions, an examination of the reported data suggests equivalent activation in PHC for scenes that did and did not explicitly depict local 3D scene structure. However, this comparison is qualitative and post hoc, and in any case a functionally defined PPA ROI was not reported. Thus, although there is a tentative suggestion that PPA activates equivalently to close-up and full scene views (contrary to what might be expected on the 3D geometry hypothesis), this evidence is preliminary at best.

An alternative hypothesis concerning pPHC function suggests that it is involved in processing relational information among objects that typically appear in specific scene contexts (Bar & Aminoff, 2003; also Bar, 2004). It is well-known that places and landmarks such as buildings activate pPHC (Maguire, Frith, Burgess, Donnett, & O'Keefe, 1998; Mellet et al., 1996). On the contextual association

hypothesis, the place-specificity observed for pPHC can be reinterpreted as context-specificity, with place comprising one type of strongly associated context. This hypothesis suggests that highly scene-diagnostic objects such as a stoves, beds, and couches might be expected to activate pPHC as well as their associated full scenes. Bar and Aminoff (2003, Experiment 1) compared strong contextual association objects without background (termed by the investigators *strong CA_I*), strong contextual association objects with background (termed *strong CA_B*), and weak contextual association objects (*Weak CA*). Most critically for current purposes, the strong contextual association objects with and without background did not produce statistically different fMRI signal change in PHC, and both differed from the weak association objects (though there was a qualitative ordering of percent signal change, with a non-significant trend for greater activation when background was present). These results suggest that objects with strong contextual associations to scenes might activate pPHC as well as full scenes. Importantly, though, pPHC showed additional activation to scenes that included 3D spatial structure over those that did not, suggesting that pPHC may be involved in analysis of both contextual associations and spatial structure.

In contrast to the results of Bar and Aminoff (2003), Epstein and Kanwisher (1998) reported the results of an fMRI study demonstrating that arrays of objects extracted from scenes did not activate PPA, whereas the empty 3D scene backgrounds from which the objects had been removed did. These results suggest that whereas spatial geometry is sufficient to produce PPA activation, scene-relevant objects are not. On the face of it, this result appears at odds with the results of Bar and Aminoff (2003). It may be, however, that the objects used by Bar and Aminoff (2003) were more diagnostic of specific scenes (and so had stronger contextual associations) than those used by Epstein and Kanwisher (1998). Given the statistical ambiguity of the Bar and Aminoff (2003) data and the inconsistency of the results across studies, an additional test of strong-association objects to equivalent objects with some background (that is, close-up scenes), as well as to full scenes, is needed. In the present study we sought to determine if strong contextually associated objects alone are sufficient to activate pPHC equivalently to full scenes, or if instead activation increases as the stimulus becomes more completely scene-like.

A second cortical region implicated in real-world scene processing is retrosplenial cortex (RSC). RSC had traditionally been associated with episodic memory and spatial navigation (Maguire, 2001), and there is some evidence that it may also play a role in analysis of emotion (Maddock, 1999). Most relevant for present purposes, RSC (particularly Brodmann's Area 30) has reciprocal connections with parahippocampal cortex, and appears to be activated in many of the same navigation tasks as pPHC. In a comprehensive review, Maguire (2001) noted that lesion studies suggest RSC (particularly right RSC and even more

particularly right BA 30) is involved in way-finding along familiar routes as well as learning new routes. Maguire also observed that functional neuroimaging studies implicate bilateral RSC (again particularly BA 30) in navigation and spatial orientation within large-scale space. However, as with pPHC, it has been suggested that these results could be reinterpreted with RSC functioning to support associations related to scene contexts (Bar & Aminoff, 2003). Supporting this hypothesis, Bar and Aminoff (2003) found that RSC was equally activated by contextually associated objects whether those objects appeared alone or in scene backgrounds, and that both of these conditions produced greater RSC activation than objects that were weakly associated with specific contexts. Bar and Aminoff concluded that perception of highly contextual objects is sufficient to produce context-specific activation in RSC. Here we provide a new test of the spatial structure versus contextual association hypotheses for RSC function using real-world scene images.

2. Present study

We investigated the function of pPHC and RSC with respect to scene perception by directly comparing, in a single experiment, fMRI activation to full views of indoor scenes, close-up views of sub-regions of scenes from the same scene category, and single highly scene-diagnostic objects from that category (Fig. 1). Both the full-view and close-up scene images contained objects that were highly diagnostic of the scene category. The Close-up Scene and diagnostic object conditions were equivalent to the

contextual association with background (CA_B) and contextual association without background (CA_I) conditions used by Bar and Aminoff (2003). We also included faces as a control condition against which to test the scene-relevant stimuli.

Prior studies have typically presented small, repeated sets of images. However, it is known that activation declines with image repetition (Grill-Spector et al., 1999; Grill-Spector & Malach, 2001), and this effect has been shown to hold in PPA (Epstein et al., 2003). Greatest sensitivity to condition differences would therefore be expected with non-repeated images. In the present study we presented subjects with a large set of non-repeating images in each condition. This method increased statistical power to detect differences across conditions, and also helped ensure that the results generalize to the population of stimulus types under investigation. Fourteen participants each saw 480 unique pictures in a block-design fMRI experiment. Because we were interested in scene perception and interpretation rather than intentional memory encoding or retrieval, participants passively viewed the images (Goh, 2004). All images in the three critical conditions (Full Scenes, Close-up Scenes, and Diagnostic Objects) came from the same basic-level scene category (kitchens; see Fig. 1). If PPA and RSC primarily process contextual scene associations, then activation to highly diagnostic objects would be expected to be as strong as activation to the diagnosed scenes. If PPA and RSC are responsible for both contextual associations and 3D scene geometry, then activation to diagnostic objects and close-up scenes may be similar (as both strongly suggest a specific scene



Fig. 1. Examples of images in the four conditions. One-hundred twenty unique Full Scenes (kitchens), Close-up Scenes (stove areas), Diagnostic Objects (stoves), and Faces were presented without repetition to each subject.

category), but activation to full scenes that include explicit 3D structure should be greater. If PPA and RSC only process local 3D structure, then no activation over baseline would be expected for close-up scenes or diagnostic objects, but full scenes explicitly depicting 3D spatial structure should produce additional activation over these other conditions. Finally, differential results across PPA and RSC would suggest that these areas support analysis of different scene properties.

3. Material and methods

3.1. Subjects

Fifteen right-handed, healthy college students from Michigan State University volunteered to participate in this study. All subjects signed the consent forms approved by the Michigan State University Institutional Review Board. Data from one subject was discarded due to lack of activation in primary visual areas suggesting lack of attention during the study. Fourteen subjects (half male and half female, mean age 21.8, range 18–32) were included in the data analysis.

3.2. Stimuli

Stimuli were 480 full-color digitized photographs selected from a variety of sources, with 120 unique pictures for each of the four conditions (Fig. 1). Faces were chosen from the AR Face Database (Martinez & Benavente, 1998). Stimuli were displayed in color on a 640 × 480 LCD monitor mounted on top of the RF head coil. The LCD subtended 12° × 16° of visual angle.

3.3. Procedure

A block-design paradigm was controlled by an IFIS-SA system (Invivo Corp., Gainesville, FL). The experiment was divided into four functional runs each lasting 8 min and 15 s. In each run, subjects were presented with 12 blocks of visual stimulation after an initial 15 s “resting” period. In each block, 10 unique pictures from one condition were presented. Within a block, each picture was presented for 2.5 s with no inter-stimulus interval. A 15 s baseline condition (a white screen with a black cross at the center) followed each block. Each condition was shown in three blocks per run. Both the order of conditions within each run and the order of pictures within a block were initially randomly determined. The four functional runs were presented to half of the subjects in a forward order and half in a reverse order.

3.4. Image acquisition

The experiment was conducted on a 3T GE Signa EXCITE scanner (GE Healthcare, Milwaukee, WI) with an 8-channel head coil. During each session, images were

first acquired for the purpose of localization, and first and higher-order shimming procedures were then carried out to improve magnetic field homogeneity (Kim, Adalsteinsson, Glover, & Spielman, 2002). To study brain function, echo planar images, starting from the most inferior regions of the brain, were then acquired with the following parameters: 34 contiguous 3-mm axial slices in an interleaved order, TE = 25 ms, TR = 2500 ms, flip angle = 80°, FOV = 22 cm, matrix size = 64 × 64, ramp sampling, and with the first four data points discarded. Each volume of slices was acquired 194 times during each of the four functional runs while subjects viewed the pictures, resulting in a total of 776 volumes of images over the course of the entire experiment. After functional data acquisition, high-resolution volumetric T_1 -weighted spoiled gradient-recalled (SPGR) images with cerebrospinal fluid suppressed were obtained to cover the whole brain with 124 1.5-mm sagittal slices, 8° flip angle and 24 cm FOV. These images were used to identify anatomical locations.

3.5. fMRI data pre-processing and analysis

All fMRI data pre-processing and analysis was conducted with AFNI software (Cox, 1996). For each subject, with the first functional image as the reference, rigid-body motion correction was done in three translational and three rotational directions. The amount of motion in these directions was estimated and then the estimations were used in data analysis. For each subject, spatial blurring with a full width half maximum of 4 mm was applied to reduce random noise (Parrish, Gitelman, LaBar, & Mesulam, 2000), and also to reduce the issue of inter-subject anatomical variation and Talairach transformation variation during group analysis. For the group analysis, all images were converted to Talairach coordinate space (Talairach & Tournoux, 1988) with an interpolation to 1 mm³ voxels.

For the data analysis of each individual subject, the reference function throughout all functional runs for each picture category was generated based on the convolution of the stimulus input and a gamma function (Cox, 1996), which was modeled as the impulse response when each picture was presented. The functional image data acquired was compared with the reference functions using the 3dDeconvolve software for multiple linear regression analysis and general linear tests (Ward, 2002). Multiple linear regressions were applied on a voxel-wise basis for t -statistic tests and to find the magnitude change when each picture condition was presented, compared to the reference functions. The equivalent BOLD percent signal change relative to the baseline state was then calculated. General linear tests were also applied on a voxel-wise basis to find the statistical significance of pair-wise comparisons for all the picture conditions. For the above analysis, in addition to applying the reference functions for the four picture conditions, MRI signal modeling also included the subject motion estimations in the three translational and the three

rotational directions, and the constant, linear and quadratic trends for each of the four functional runs.

Monte Carlo simulation of the effect of matrix and voxel sizes of the imaging volume, spatial correlation of voxels, voxel intensity thresholding, masking and cluster identification was applied to estimate overall statistical significance with respect to the whole brain (Ward, 2000).

3.6. Parahippocampal place area (PPA) and retrosplenial cortex (RSC) region of interest (ROI) analyses

In Talairach coordinate space, a PPA ROI was identified in each hemisphere for each subject. Regions showing preferential activation to Full Scenes over Diagnostic Objects (a 14-subject average whole-brain corrected p value $<1.4 \times 10^{-3}$) in the parahippocampal gyrus were defined as the PPA. The whole-brain corrected p values were the results of the following active voxel selection criteria: a voxel-wise p value $<10^{-4}$ based on the t test and active voxels that were nearest-neighbor and within a cluster size of 107 mm^3 . This contrast resulted in clear PPA ROIs in posterior parahippocampal cortex in each hemisphere for each subject. We computed the average BOLD percent signal change for all voxels in the PPA ROI individually for each subject as a function of condition and hemisphere. These data were entered into a Condition (Full Scenes, Close-up Scenes, Diagnostic Objects, and Faces) by Hemisphere (left, right) ANOVA. To minimize any bias resulting from the Full Scene condition, an alternative PPA ROI analysis was conducted in which PPA was defined as preferential activation of scene and object over face (14-subject average whole-brain corrected p value $<1.4 \times 10^{-3}$) in the parahippocampal gyrus based on the contrast [(Full Scenes + Close-up Scenes + Diagnostic Objects) – 3 × Faces].

We also conducted a subject analysis for retrosplenial cortex. An analogous Condition by Hemisphere ANOVA was calculated for the mean percent signal change by subject for all voxels in the anatomically-defined (BA 29 and 30) RSC ROI.

3.7. Whole brain analysis

After the percent signal change was estimated with respect to each picture condition for each subject, an ANOVA was performed over the fourteen subject data sets for group analysis with a mixed-effect two-factor model. Picture condition (four levels) was the first factor and was modeled to provide a fixed effect. Subject was the second factor and was modeled as a random effect. The ANOVA results were used to extract the activated voxels for all pair-wise condition contrasts (voxel-based p value $<10^{-3}$ and whole-brain corrected p value $<1.9 \times 10^{-3}$). The active voxel selection criteria required that the voxels were nearest-neighbor and within a cluster size of

177 mm^3 . Based on application of these criteria to the whole brain, the voxel-based p value $<1 \times 10^{-3}$ was corrected to be an equivalent of whole-brain corrected p value $<1.9 \times 10^{-3}$.

4. Results

4.1. PPA ROI analysis

All 14 subjects showed well-defined PPA ROIs, with an average right PPA ROI volume of 5307 mm^3 (range 679 – 9957 mm^3), and average left PPA ROI volume of 4641 mm^3 (range 458 – 9177 mm^3). The group analysis also showed that the PPA was more extensive in the right than the left hemisphere (4133 versus 2483 mm^3), with region centroids of $24, -40, -4$ and $-22, -42, -4$ (Talairach coordinates) for the right and left hemispheres, respectively, consistent with previous reports (Fig. 2). The Condition by Hemisphere ANOVA for PPA revealed significant effects of Condition, $F(3, 39) = 72.30, p < .001, \eta^2 = .848$, and Hemisphere, $F(1, 13) = 4.99, p \leq .05, \eta^2 = .277$, which reflected greater right than left PPA activation. The Condition × Hemisphere interaction did not reach significance, $p > .30$ (Fig. 3). There was a clear ordering of PPA activation as a function of condition, with greatest activation for Full Scenes, and progressively less for Close-up Scenes followed by Diagnostic Objects, with no discernible activation for faces. Subsequent t -tests confirmed these observations, with all ordered comparisons for each neighboring Condition and Hemisphere reaching statistical significance following Bonferroni correction. PPA activation was stronger for Full-Scenes compared to Close-up Scenes for both left and right hemispheres, $t(13) = 10.54, p < .001$ and $t(13) = 9.61, p < .001$, respectively. Close-up Scenes activated PPA more strongly than Diagnostic Objects: L PPA $t(13) = 4.50; p = .001$, R PPA $t(13) = 5.19, p < .001$. Finally, Diagnostic Objects activated PPA more strongly than faces: L PPA $t(13) = 3.33, p = .005$; R PPA $t(13) = 2.59, p < .02$.

The alternative PPA ROI analysis defined as the scene and object versus face conditions showed the same order of PPA activation as a function of condition as above, with greatest activation for Full Scenes, and progressively less for Close-up Scenes followed by Diagnostic Objects, with no discernible activation for Faces. The Condition by Hemisphere ANOVA for PPA revealed significant effects of Condition, $F(3, 39) = 139.56, p < .001, \eta^2 = .915$, with all post hoc contrasts between conditions showing a significant difference from each other in the step-wise fashion stated above ($ps < .04$). A main effect for Hemisphere, $F(1, 13) = 6.96, p = .02, \eta^2 = .349$ indicating greater activation in the right than left hemisphere was qualified by a significant Condition × Hemisphere interaction, $F(3, 39) = 4.15, p < .04, \eta^2 = .242$. Bonferroni post hoc comparisons indicated that the right hemisphere advantage was present only for Full Scenes ($p < .05$), but not the other three conditions.

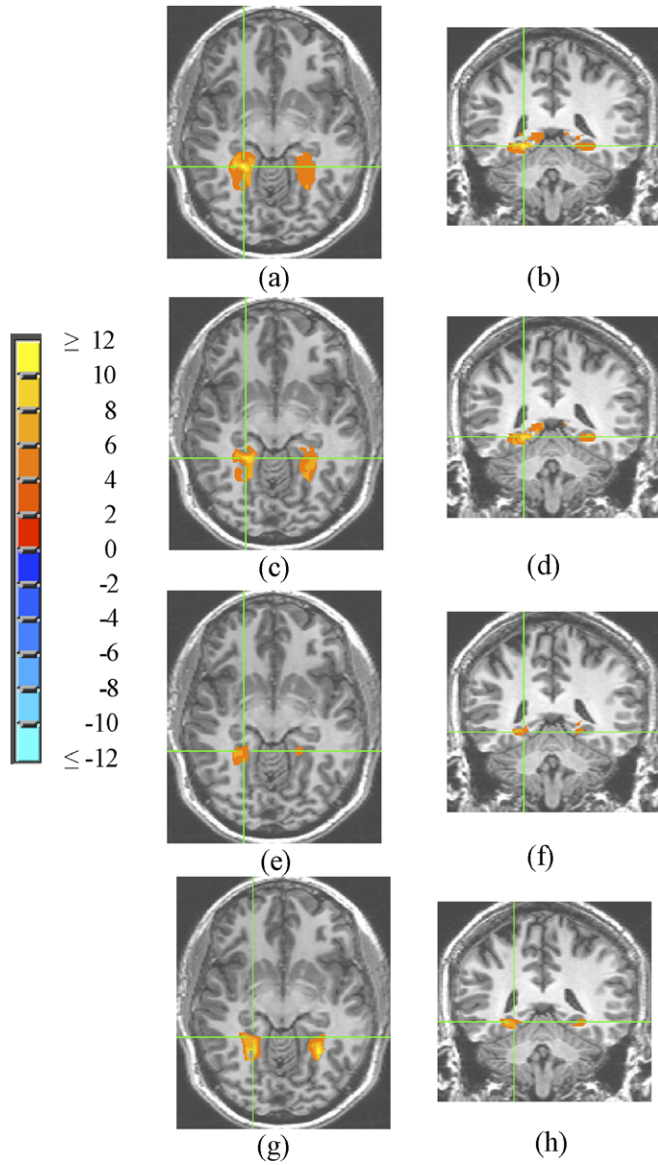


Fig. 2. Spatial extent of activation in the PPA ROI from the group analysis (*t* value scale on left). Images (a) and (b) show PPA ROI definition (Full Scenes versus Diagnostic Objects contrast) in axial and coronal views. The green crosshair indicates peak *t* value in the right PPA from this ROI definition contrast (Talairach 23, –41, –5). Images (c) and (d) show the results of the Full Scenes versus Close-up Scenes contrast, (e) and (f) show the Close-up Scenes versus diagnostic objects contrast, and (g) and (h) show the Diagnostic Objects versus Faces contrast.

4.2. RSC ROI analysis

In an analogous Condition × Hemisphere ANOVA on the RSC ROI we found a significant main effect of condition, $F(3,39) = 16.91, p < .001, \eta^2 = .565$ (Fig. 4). Neither the main effect of Hemisphere nor the Condition × Hemisphere interaction reached significance, $F_s < 1$. Post hoc contrasts for Condition indicated that Full Scenes more strongly activated RSC than any of the other three conditions, Close-up Scenes, Objects, or Faces, $p_s < .005$ (Bonferroni adjusted). After Bonferroni correction, none of

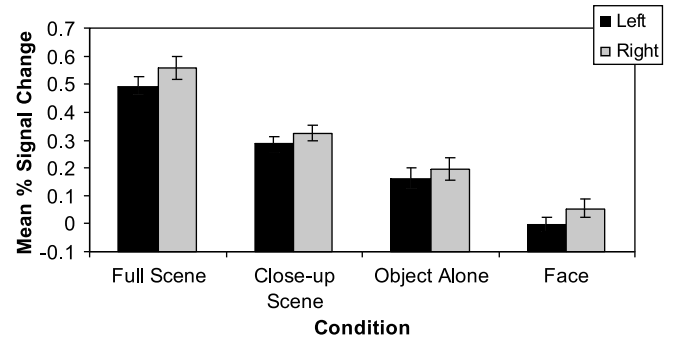


Fig. 3. Mean percent signal change (and SEM) across all voxels in the PPA ROI for 14 subjects as a function of Image Condition and Hemisphere.

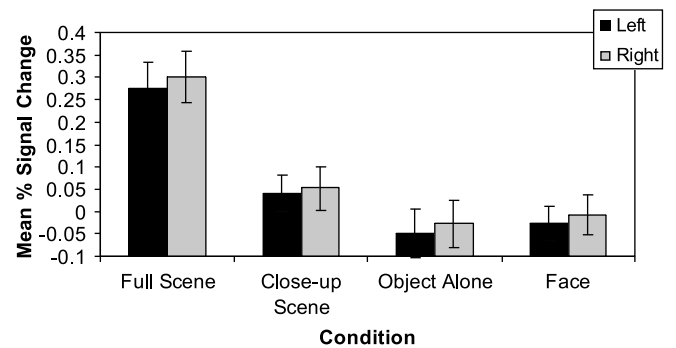


Fig. 4. Mean percent signal change (and SEM) across all voxels in the RSC ROI for 14 subjects as a function of Image Condition and Hemisphere.

the other three conditions significantly differed from each other. However, there was a trend for Close-up Scenes to elicit greater activation than diagnostic objects ($p < .05$, uncorrected).

4.3. Whole-brain group analysis

Voxel-wise group analyses were conducted contrasting Full Scene versus Close-up Scene, Full Scene versus Diagnostic Object, Close-up Scene versus Diagnostic Object, and Diagnostic Object versus Faces. For the purposes of the whole-brain group analysis, regions showing preferential activation of Full Scenes (kitchens) over Diagnostic Objects (stoves) from the group data in the parahippocampal gyrus were defined as the PPA. These analyses confirmed the PPA and RSC ROI results and further revealed that the RSC activation was limited to BA 30 (see Table 1). In addition to pPHC and RSC, a number of additional regions also demonstrated differential activation as a function of scene condition. A large number of voxels spanning the anterior-posterior length of the lingual gyrus showed greater activation to scenes compared to objects. While the bulk of this activated region lies in posterior general visual processing areas (BA 18 and 19) and revealed a preference for full scenes compared to all other conditions, the anterior portion of the lingual gyrus just

Table 1
Group analysis results for whole-brain analysis for PHC and RSC

Region	Full-Scene > Close-up Scene			Close-up Scene > Object			Object > Face					
	Volume (mm ³)	Coordinate x, y, z	FS % Signal Change (SD)	CS % signal change (SD)	Volume (mm ³)	Coordinate x, y, z	CS % signal change (SD)	O % signal change (SD)	Volume (mm ³)	Coordinate x, y, z	O % signal change (SD)	F % signal change (SD)
<i>PHC</i>												
R	3540	21, -42, -4	0.60 (0.36)	0.34 (0.31)	668	29, -43, -4	0.59 (0.27)	0.39 (0.22)	2100	26, -52, -4	0.34 (0.30)	0.15 (0.22)
L	2275	-25, -48, -4	0.58 (0.32)	0.34 (0.29)	196	-19, -41, 0	0.23 (0.15)	0.07 (0.11)	1644	-26, -51, -5	0.38 (0.26)	0.06 (0.15)
<i>RSC</i>												
BA 30												
R	1079	(18, -40, -6) & (9, -55, 9)	0.34 (0.12)	0.04 (0.10)								
L	867	-14, -57, 14	0.32 (0.14)	0.03 (0.10)	153	-18, -54, 10	0.14 (0.05)	0.00 (0.04)				
BA 29												
R	102	8, -50, 6	0.53 (0.21)	0.13 (0.09)								

For each contrast, cluster size, the Talairach coordinates for the peak *t*-values and mean percent signal change for the cluster are reported. RSC was analyzed separately for BA 29 and BA 30. Mean percent signal change is not reported for any region that did not contain voxels exceeding the statistical significance threshold. FS, Full Scene; CS, Close-up Scene; O, Object; F, Face; RSC, retrosplenial cortex.

caudal to the PHC appears to be more functionally similar to the PPA. As seen in Table 2, a portion of lingual gyrus exhibited the same pattern observed in PHC, with a strong preference for Full Scenes over Close-up Scenes and Close-up Scenes over Diagnostic Objects. These voxels all lie in anterior lingual gyrus and are contiguous with the PHC cluster showing significant activation for this contrast. These data are in keeping with previous work suggesting that the functionally-defined PPA sometimes extends slightly posteriorly from PHC into the lingual gyrus (Epstein, 2005).

Full Scenes also most strongly activated posterior cingulate cortex. In addition, a small cluster in left BA 30 (RSC) also showed greater activation for Close-up Scenes compared to Diagnostic Objects. Precuneus activation was present for all types of stimuli, including Faces versus control. However, activation was greatest for Full Scenes compared to all other conditions. Other visual association areas including the cuneus, fusiform gyrus, and middle occipital gyrus also showed stronger, but not differential, activation to the three scene conditions compared to faces.

5. Discussion

Understanding how scene-specific representations and processes are implemented in the brain is a central issue in visual cognition and cognitive neuroscience. The current study explored specific hypotheses about the cognitive functions that purported cortical scene-processing areas participate in. Prior research has suggested that at least two cortical regions, pPHC (and particularly PPA) and RSC, play an important role in scene analysis, but the specific functions of these regions remains unclear. The purpose of the present study was to investigate two competing hypotheses concerning the function of PPA and RSC in scene analysis, the 3D spatial geometry hypothesis and the contextual association hypothesis.

Scene images differ in the degree to which the 3D structure of the environment is explicitly available. The spatial geometry hypothesis predicts that scene-specific activation should increase as more 3D geometry becomes perceptually explicit in an image. In a full view of a room, the 3D structure is apparent from the larger-scale background elements and surfaces such as walls, floor, and ceiling. In a close-up view of a limited area of a room, objects and surfaces remain visible, but the 3D structure of the room is no longer directly apparent. If scene regions preferentially process 3D structure, then we would expect more activation to full views than to close-up views. If instead scene regions preferentially process scenes regardless of the explicitness of 3D structure, then activation should be equivalent in these conditions.

The contextual association hypothesis suggests that a primary function of scene areas is to process scene-relevant relationships among objects and their context (Bar, 2004). Behavioral and computational research demonstrate that scene context and gist can be generated both by an analysis

Table 2
Group analysis results for whole-brain analysis for additional brain regions showing significant differences between conditions

Region	Full-Scene > Close-up Scene		Close-up Scene > Object		Full-Scene > Object	
	Volume (mm ³)	Coordinate x, y, z	Volume (mm ³)	Coordinate x, y, z	Volume (mm ³)	Coordinate x, y, z
<i>Lingual gyrus</i>						
R	9812	7, -85, 5	2551	8, -79, -5	8129	7, -77, -5
L	8258	-25, -65, -3	949	-7, -82, -3	7066	(-13, -89, -3) & (-24, -54, 0)
<i>Fusiform gyrus</i>						
R	563	29, -39, -11				
L	562	-23, -67, -7			338	-23, -67, -7
<i>Posterior cingulate</i>						
R	2036	10, -55, 10			2034	19, -53, 12
L	1940	-14, -59, 15	335	-18, -54, 10	1994	-15, -60, 15
<i>Middle occipital gyrus</i>						
R	2312	22, -90, 7			1964	20, -87, 15
L	1522	-28, -87, 0			1151	-18, -95, 6
<i>Precuneus</i>						
R	277	(9, -58, 57) & (18, -63, 25)				

For each contrast, cluster volume and Talairach coordinates for the peak *t*-value for the cluster are reported. Clusters not exceeding statistical and clustering thresholds are not listed.

of scene-specific visual information (Torralba, 2003) and by highly scene-diagnostic objects (Davenport & Potter, 2004; Henderson & Hollingworth, 1999). Therefore, if cortical scene regions primarily process scene-relevant contextual associations, then equal activation would be expected for scene-diagnostic objects and to their diagnosed scenes (Bar & Aminoff, 2003).

In the present study, we contrasted three critical conditions: Full Scenes depicting an indoor environment from a normal viewing distance, Close-up Scenes depicting a more limited region of the same environment from a nearer vantage point, and single diagnostic objects that were highly associated with their scene environment. These scene conditions were compared with a Face control condition. With respect to the PPA, results of both region-of-interest and whole-brain analyses converged on the conclusion that PPA activation is greatest for full views, with progressively decreasing activation for close-up views of scene regions, followed by single scene-diagnostic objects, with no activation for faces. These results show that PPA preferentially responds to scenes that explicitly depict the 3D geometrical structure of the local visual environment than those that do not (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998; Epstein et al., 2003). The results also show that PPA preferentially responds to scenes over highly diagnostic objects associated with those scenes. These results are consistent with the hypothesis that PPA is a scene processing region involved in computations related to analyzing the local spatial environment. Furthermore, activation in PPA was greater for scenes that included more spatial structure than for close-up scenes, even though the latter included strongly scene-diagnostic objects. These results suggest that PPA processes more than scene-relevant contextual associations. At the same time, clear PPA activation was observed for scene-diagnostic

objects over faces (which did not differ from baseline), suggesting that PPA may be involved to some degree in non-spatial relational processing as well (Bar & Aminoff, 2003).

In contrast to PPA, RSC did not show a graded response as the scene image more explicitly included 3D geometry. Instead, activation beyond that observed for faces was restricted to the full scene condition. These results are not compatible with the hypothesis that RSC is involved in associative processing, since if it were, some activation to both scene-diagnostic objects and close-up scenes (which included scene-diagnostic objects) would be expected over faces. At the same time, there was a (non-significant) tendency toward greater activation for the close-up scenes than the diagnostic objects. These results are mildly compatible with the hypothesis that RSC is also involved in processing the geometric structure of scenes, a result consistent with the general idea that RSC participates in neural circuits supporting spatial navigation. Interestingly, the majority of the activity in RSC was restricted to BA 30. This latter result is compatible with the findings of past studies suggesting that only a smaller area of RSC is involved in scene processing (Maguire, 2001).

Finally, we consider two alternative explanations for the activation advantage in PPA and BA 30 for scenes with explicit 3D structure. First, such scenes might preferentially attract or engage attention. Although this is a logical possibility, without a theoretically motivated *a priori* explanation for why scenes with 3D geometry preferentially engage attention, this account has no explanatory power. We therefore apply Occam's Razor and propose that it is the 3D structure itself that is the source of the effect. Second scenes with explicit 3D structure might contain more discrete objects, and the number of individuated objects (or the number of spatial relationships among objects) might be the source of the differential activation. Contrary to this

hypothesis, however, in a direct comparison of PPA response to empty rooms versus furnished rooms versus arrays of objects, Epstein and Kanwisher (1998) found that activation was equivalent to empty and furnished rooms, whereas activation to both types of rooms was over twice as great as to the object arrays. At the same time, activation to object arrays was equivalent to single objects. From these results it would appear that the number of objects or spatial relations among objects is not the critical determinant of activation in posterior parahippocampal cortex. Instead, it appears that 3D structure directly drives activation in these regions.

6. Conclusion

In conclusion, the present results converge with prior studies in suggesting that at least two cortical areas, pPHC (specifically PPA) and RSC, are preferentially involved in the processing of real-world scenes over complex and meaningful visual stimuli including faces and objects. The results further demonstrate that PPA shows a preference for images that include scene background over scene-diagnostic objects alone, and greatest preference for scenes in which the geometric 3D structure of the space is explicitly depicted. These results strongly suggest that PPA preferentially processes the 3D geometrical structure of the local visual environment. RSC similarly shows strong preference for scene images in which 3D structure is explicit, but does not appear otherwise to prefer scene images or single diagnostic objects. One interpretation of these latter results is that RSC is even more specifically tuned to spatial information related to 3D geometric structure.

Acknowledgments

This research was supported by grants from the MSU Foundation to John M. Henderson, and from the MSU Intramural Research Grant Program and NIMH (MH071275-01) to Christine L. Larson. We also thank Michigan State University's Department of Radiology, Office of the Vice President for Research and Graduate Studies, and Core Analytical Fund for their support. We thank David McFarlane, Jie Huang, Jeremy Grounds, Shawn Katterman, Stephen Kemsley, and Jordan Robinson for their assistance on the project.

References

- Aguirre, G., & D'Esposito, M. (1997). Environmental knowledge is subserved by separate dorsal/ventral areas. *Journal of Neuroscience*, *17*, 2512–2518.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.
- Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*, 347–358.
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance images. *Computers and Biomedical Research*, *29*, 162–173.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*, 559–564.
- Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*, 954–978.
- Epstein, R., Graham, K., & Downing, P. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, *37*, 865–876.
- Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, *23*, 115–125.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Goh, J. O. S. (2004). Cortical areas involved in object, background, and object-background processing revealed with functional magnetic resonance adaptation. *The Journal of Neuroscience*, *24*, 10223–10228.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, 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.
- Henderson, J. M., Larson, C. L., & Zhu, D. C. (2007). Cortical activation to indoor versus outdoor scenes: An fMRI study. *Experimental Brain Research*, *179*, 75–84.
- Henderson, J. M., & Ferreira, F. (Eds.). (2004). *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
- Henderson, J., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.
- Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 604–610.
- Kim, D., Adalsteinsson, E., Glover, G., & Spielman, D. (2002). Regularized higher-order in vivo shimming. *Magnetic Resonance in Medicine*, *48*, 715–722.
- Kohler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus*, *12*, 718–723.
- Li, F. F., Van Rullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 9596–9601.
- Maddock, R. J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends in Neuroscience*, *22*, 310–316.
- Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology*, *42*, 225–238.
- Maguire, E. A., Frith, C. D., Burgess, N., Donnett, J. G., & O'Keefe, J. (1998). Knowing where things are: Parahippocampal involvement in encoding object locations in virtual large-scale space. *Journal of Cognitive Neuroscience*, *10*, 61–76.
- Maguire, E. A., Frith, C. D., & Cipolotti, L. (2001). Distinct neural systems for the encoding and recognition of topography and faces. *Neuroimage*, *13*, 743–750.
- Martinez, A. M., & Benavente, R. (1998). The AR face database. CVC Tech. Report #24.
- Mellet, E., Tzourio, N., Crivello, F., Joliot, M., Denis, M., & Mazoyer, B. (1996). Functional anatomy of spatial mental imagery generated from verbal instructions. *The Journal of Neuroscience*, *16*, 6504–6512.
- Parrish, T., Gitelman, D., LaBar, K., & Mesulam, M. (2000). Impact of signal-to-noise on functional MRI. *Magnetic Resonance in Medicine*, *44*, 925–932.

- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509–522.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5, 195–200.
- Talairach, J., & Tournoux, P. (1988). *Co-Planar stereotaxic atlas of the human brain: 3D proportional system*. New York: Georg Thieme Verlag.
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522.
- Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America, A, Optics, Image Science and Vision*, 20, 1407–1418.
- Torralba, A., Oliva, A., Castelhana, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766–786.
- Ward, B. (2000). *Simultaneous inference for fMRI data*. Milwaukee, WI: Biophysics Research Institute, Medical College of Wisconsin.
- Ward, B. (2002). *Deconvolution analysis of fMRI time series data*. Milwaukee, WI: Biophysics Research Institute, Medical College of Wisconsin.