

Differential Parahippocampal and Retrosplenial Involvement in Three Types of Visual Scene Recognition

Russell A. Epstein and J. Stephen Higgins

Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania, PA, USA

Human observers can quickly and accurately interpret the meaning of complex visual scenes. The neural mechanisms underlying this ability are largely unexplored. We used functional magnetic resonance imaging to measure cortical activity while subjects identified briefly presented scenes as specific familiar locations (“Houston Hall”), general place categories (“kitchen”), or general situational categories (“party”). Scene-responsive voxels in the parahippocampal place area (PPA) and retrosplenial cortex (RSC) were highly sensitive to recognition level when identifying scenes, responding more strongly during location identification than during place category or situation identification. In contrast, the superior temporal sulcus, cingulate sulcus, and supermarginal gyrus displayed the opposite pattern, responding more strongly during place category and situation identification. Consideration of results from 4 experiments suggests that the PPA represents the visuospatial structure of individual scenes, whereas RSC supports processes that allow scenes to be localized within a larger extended environment. These results suggest that different scene identification tasks tap distinct cortical networks. In particular, we hypothesize that the PPA and RSC are critically involved in the identification of specific locations but play a less central role in other scene recognition tasks.

Keywords: fMRI, object recognition, parahippocampal place area, place recognition, spatial memory, visual system

Introduction

A large body of research has examined the neural and psychological basis of object recognition (Biederman 1987; Tanaka 1993; Logothetis and Sheinberg 1996). Outside of the laboratory, however, discrete objects such as faces, bottles, and shoes are never encountered in isolation; they are always part of a larger, surrounding scene, which can be encoded and recognized in its own right (Intraub 1997; Henderson and Hollingworth 1999; Oliva and Torralba 2001; Chun 2003; Torralba and Oliva 2003). Behavioral experiments have demonstrated that subjects can identify scene meaning or “gist” very rapidly (Biederman and others 1974; Potter 1975; Schyns and Oliva 1994; Renninger and Malik 2004; Maljkovic and Martini 2005; Rousselet and others 2005), suggesting that the visual system might contain specialized machinery for scene comprehension. Although regions that respond preferentially to real-world visual scenes have been identified (Epstein and Kanwisher 1998) and the coordinate frames of scene representations within these regions explored (Epstein and others 2003, 2005), the neural bases of the mechanisms underlying scene recognition have not been studied in depth. For example, it is not known whether there is a single cortical network involved in scene recognition or whether different scene

recognition tasks might tap substantially different neural systems. This state of affairs contrasts notably to the object recognition literature, where questions about possible distinct neural systems for category-specific (Puce and others 1996; McCarthy and others 1997; Haxby and others 2000; Kanwisher 2004; Downing and others 2005) or task-specific (Diamond and Carey 1986; Gauthier and others 1999; Gauthier 2000) recognition processes have been extensively addressed.

As with objects, there are at least 2 different levels at which a scene can be identified (Rosch and others 1976; Tversky and Hemenway 1983). Consider the case in which the scene in question is a view of a real-world location, encountered either as a photographic image or by viewing the location directly. A scene of this type can either be identified as an exemplar of a general place category (e.g. “a store”), or as a specific place with a specific location in the world (e.g. “the Penn Bookstore at 36th and Walnut St.”). Both types of information are useful, but in different ways. Knowing what kind of place you are in is critical for knowing how to act in the place, and can also be a useful cue to facilitate the recognition of the objects within the scene (Biederman 1972; Bar 2004; Davenport and Potter 2004; but see Hollingworth and Henderson 1998). Knowing what specific place you are in tells you where you are in the world, allowing orientation within the larger spatial environment. Although information about place category and place identity are clearly linked together at some point, it is possible that the processing streams that are used to extract these 2 kinds of information from visual input might be at least partially distinct. For example, these different recognition tasks may require representation of different aspects of the visual scene, such as component object identity for place categorization and spatial layout for location identification. Furthermore, category identification and location identification may require activation of anatomically distinct long-term memory codes, one for the categorical and one for the topographical qualities of the local scene. The first goal of the current study was to address this issue by determining whether the neural systems that mediate place recognition at the basic level (“What kind of place is this”) are similar to those that mediate place recognition at the exemplar level (“What specific place is this?”).

In the preceding example, the primary aspect of the scene is its identity as a place. However, some scenes are defined primarily by what is happening in the scene (“a party,” or “a picnic”) rather than by the enduring features of the background environment. It is unclear to what extent recognition of these “situational” scenes taps the same mechanisms involved in place recognition. In both cases, identification of the scene requires analysis of more than a single element or object, which might suggest the use of common mechanisms for visual integration. However, the defining aspects of places tend to be large fixed

elements (such as walls, furniture, and major appliances), whereas the defining aspects of situations tend to be smaller dynamic elements (such as people, or small movable objects). Insofar as previous work has suggested different neural systems involved in processing people versus places (Kanwisher 2004), living things versus nonliving things (Martin and others 1996), and static versus dynamic aspects of objects (Haxby and others 2000; Beauchamp and others 2002) one might expect that the representations used for recognition of places and recognition of situations would be distinct. A second goal of this study was to address this issue by determining the extent to which the neural systems that mediate recognition of spatial scenes (i.e. places) are also engaged during recognition of social scenes (i.e. situations).

These questions were investigated by scanning subjects with functional magnetic resonance imaging (fMRI) while they performed 3 different scene recognition tasks. In the *location identification* task, subjects identified specific locations around a familiar college campus (e.g. "Franklin Building—36th and Walnut," "Upper Quad Gate—37th and Spruce"); in the *category identification* task, they identified places at the categorical level (e.g. "kitchen," "parking lot"); in the *situation identification* task, they identified different kinds of situations involving people (e.g. "date," "protest"). We reasoned that comparison of neural activity between the location identification and category identification tasks would allow us to identify cortical regions that were differentially involved in recognition of places at the specific/exemplar level (location identification task) and the basic level (category identification task). Similarly, we reasoned that comparison of neural activity between the category identification and situation identification tasks would allow us to identify cortical regions that were differentially involved in recognition of places (category identification task) and recognition of situations (situation identification task). Note that both the category and situation identification tasks require recognition at the basic level; thus, the 2 questions of interest are orthogonal to each other. Although it would be theoretically possible to include a task that required recognition of situations at the specific level ("my 35th birthday party"), this case was not examined in the current experiment.

Our studies focused in particular on 2 regions that have been previously identified as responding preferentially to visual scenes: the parahippocampal place area (PPA; Epstein and Kanwisher 1998) and retrosplenial cortex (RSC; O'Craven and Kanwisher 2000). The scene-specific response in these regions suggests that they encode information that is available from whole scenes but not from images of decontextualized objects. We have argued that these regions may extract spatial layout information from scenes that is useful for identification of specific locations (Epstein 2005). This claim is consistent with neuropsychological evidence (Bohbot and others 1998; Aguirre and D'Esposito 1999; Epstein and others 2001) and with results from neuroimaging studies, which have consistently found activity in PPA and RSC during spatial navigation and spatial memory tasks (Aguirre and others 1996; Maguire and others 1998; Burgess and others 2001; Ino and others 2002; Shelton and Gabrieli 2002; Rosenbaum and others 2004). However, these results do not preclude the possibility that PPA and RSC representations might be equally useful for identifying places at the basic level (Steeves and others 2004) or for recognition of situations. Indeed, it has been proposed that one of the primary functions of the PPA and RSC might be the classification of

scenes into generic contexts in order to facilitate object recognition (Bar and Aminoff 2003).

We present results from 4 experiments. To anticipate, our data indicate a surprising sensitivity of the PPA and RSC to recognition task. In particular, these regions responded most strongly during location identification, implicating them in the mediation of representations that are useful for distinguishing between different, individual real-world places. Furthermore, the results indicate that these regions may play distinct but complementary roles in location identification.

Materials and Methods

Subjects

Twenty-eight healthy right-handed volunteers with normal or corrected-to-normal vision were recruited from the University of Pennsylvania community and gave written informed consent according to procedures approved by the University of Pennsylvania institutional review board. Five subjects were run in experiment 1, 8 in experiment 2, 7 in experiment 3, and 8 in experiment 4. All volunteers were highly familiar with the Penn campus (average length of experience 3.7 ± 1.3 years).

MRI Acquisition

Scanning was performed at the Hospital of the University of Pennsylvania on a 3-Tesla Siemens Trio equipped with a Siemens body coil and a 4-channel head coil. T_2^* -weighted images sensitive to blood oxygenation level-dependent contrasts were acquired using a gradient-echo echoplanar pulse sequence (repetition time [TR] = 2000 ms, echo time [TE] = 30 ms, matrix size = 64×64 , voxel size = $3 \times 3 \times 3$ mm, 33 axial slices). Structural T_1 -weighted images for anatomical localization were acquired using a 3D MPRAGE pulse sequence (TR = 1620 ms, TE = 3 ms, interval time [TI] = 950 ms, voxel size = $0.9766 \times 0.9766 \times 1$ mm, matrix size = $192 \times 256 \times 160$). 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.

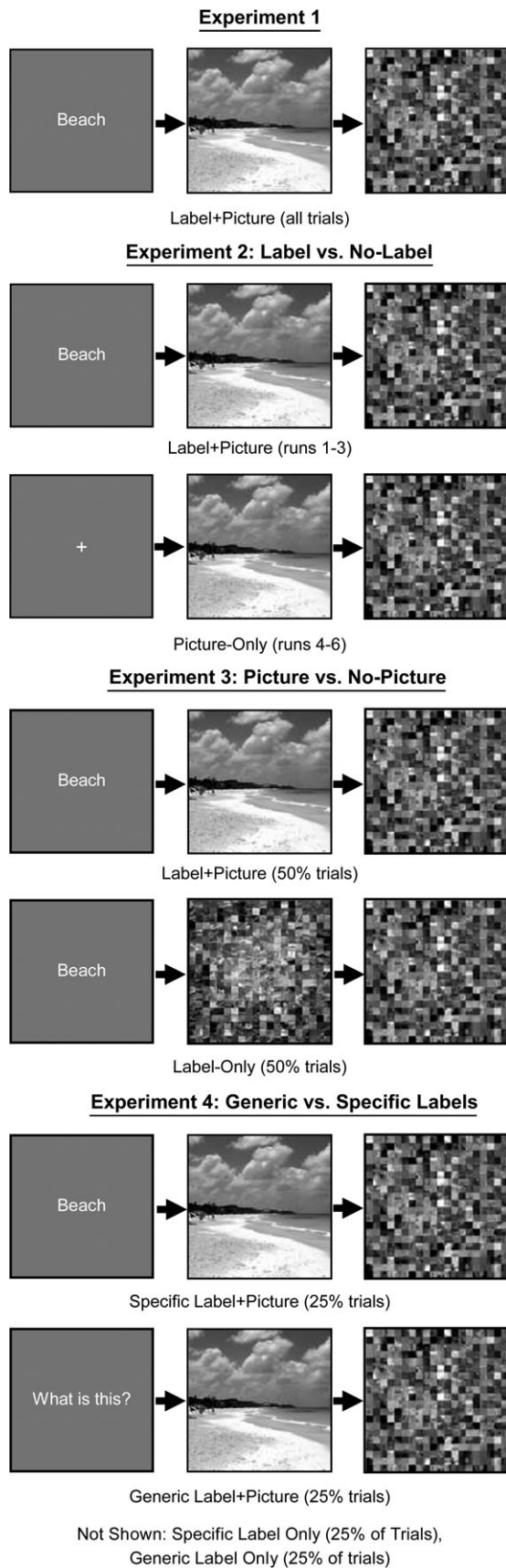
Experimental Procedure

Overview

All 4 experiments used variants of the same basic procedure, which we will describe here along with the logic behind the different variants (see Fig. 1 for illustration). Detailed information about each experiment is provided in the next section.

The basic experimental procedure was as follows. Each fMRI scan was divided up into a number of trials. In each trial, subjects saw a verbal label followed by a briefly presented and masked photograph and used a button box to indicate whether or not the label matched the photograph (Fig. 1, top). Three different recognition tasks were defined by the use of appropriate labels and photographs in different trials (Fig. 2). Specifically, in *location identification* trials, stimuli were names and photographs of familiar locations around the Penn campus; in *category identification* trials, stimuli were names and photographs of scenes that were easily categorizable as different kinds of places; in *situation identification* trials, stimuli were names and photographs of scenes that were easily categorizable as depicting different kinds of situations, usually involving people. All trials in experiment 1 followed this basic procedure, which allowed us to test whether the PPA and RSC (and other cortical regions) were sensitive to identification task during scene recognition.

Experiments 2–4 attempted to break down the basic procedure of experiment 1 into its components. In particular, we reasoned that fMRI response differences between the location, category, and situation identification trials could either be caused by differences between the image sets used in each condition (in which case removing the labels would have no effect on the pattern of results) or differences between the processes engendered by the labels (in which case removing the images might not eliminate the differences). Following this logic,



experiment 2 tested whether response differences between the identification tasks require the presence of the labels, by presenting images both with and without preceding labels (label + picture vs. picture-only trials). Similarly, experiment 3 tested whether response differences between identification tasks require the presence of the photographs, by presenting labels both with and without succeeding photographs (label + picture vs. label-only trials). Experiment 4 contrasted the effects of 2 different types of labels: specific labels, which described a specific scene and by doing so implicitly indicated the identification task (e.g. “beach,” “Penn Bookstore”), and generic labels, which simply indicated the identification task without indicating what the image might be (e.g. “what is this” for category identification, and “where is this” for location identification). This comparison allowed us to determine whether labels affected neural response by cueing retrieval of information about specific scenes, or by generically cueing the identification task.

Note that one aspect of our design was the use of different image sets for the 3 identification tasks. In theory, it should be possible to acquire images that can be simultaneously identified in terms of location, category, and situation. For example, an interior shot of the Penn Bookstore might be identifiable as “the Penn bookstore,” “a store,” or “people shopping.” In practical terms, however, it is challenging to acquire a large set of images of locations that are familiar to a heterogeneous group of subjects and also encompass a wide variety of categories and situations. For example, in the current experiment, most of our original images were outdoor campus shots and thus depicted a relatively restricted set of place categories (e.g. campus, storefront, stadium). Similarly, many of our category images were of indoor locations, which are hard to equalize for familiarity (as subjects tend to be familiar with different bedrooms, kitchens and the like). Because of these concerns, we chose to collect image sets that included a wide variety of locations, categories, and situations, and to perform control experiments to ensure that differences between these conditions could not be attributed to differences between the image sets. In particular, experiment 2 examined the response to the different images when they were not presented in the context of distinct identification tasks, allowing us to isolate effects attributable to physical differences between the image sets. Similarly, experiment 3 examined the response on trials in which no images were presented, allowing us to identify effects that cannot possibly be attributed to physical differences between the image sets.

We now describe the 4 experiments in detail.

Experiment 1

Scan sessions consisted of 6 experimental scans followed by 2 functional localizer scans and 1 scan used to determine a subject-specific hemodynamic response function (HRF). Experimental scans were 7 m 36 s long and divided into 72 4-s long stimulus trials randomly interleaved with 72 2-s long null trials and 12-s fixation periods at the beginning and end of the scan. Stimulus trials began with the presentation of a verbal label for 1 s, followed after a 500-ms blank interval by a grayscale photograph (30 or 80 ms) and a mask composed of 2 overlaid spatially scrambled scenes (400 ms). A fixation cross then appeared on the screen and remained there until the beginning of the next trial. During null trials, the fixation cross was present for 2 s and subjects made no response (thus effectively jittering the interval between stimulus trials). Relatively brief presentation times were used to force subjects to attend to the information most relevant to the recognition task required on each trial, thus maximizing the probability of observing differences between the 3 identification tasks. Two image presentation times were used to accommodate the possibility of individual differences in behavioral performance; however, this manipulation was incidental to our main hypotheses.

Figure 1. Procedure for all 4 experiments. In label + picture trials, subjects viewed the verbal label for 1000 ms, followed 500 ms later by a briefly presented (30 or 80 ms) and masked (400 ms) image, which matched the label on half the trials. Picture-only trials were identical except for the absence of the initial label. Label-only trials were identical except the coherent image was replaced by a mask. Labels in experiment 4 could either indicate specific items (“beach”) or generically cue the identification task (“what is this”).

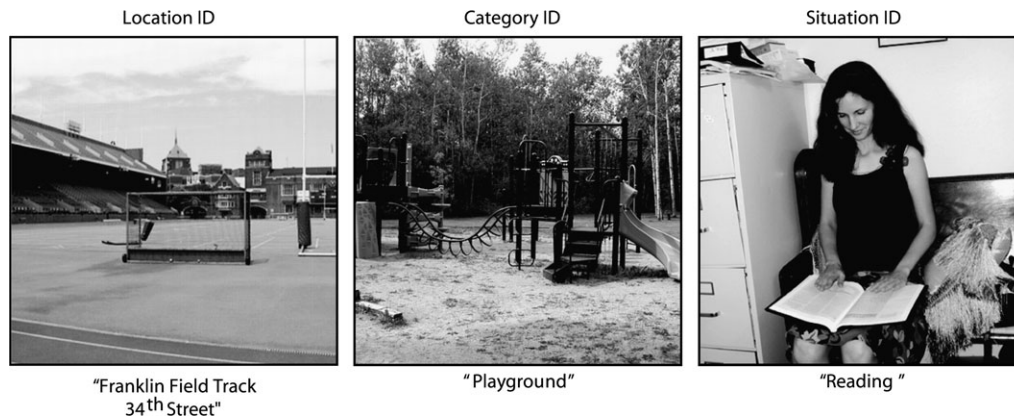


Figure 2. Examples of stimuli used (Franklin Field, Playground) or similar to those used (Reading) in the 3 identification tasks. Scenes presented in location identification and category identification trials contained few or no people, whereas almost all scenes presented in situation identification trials contained people as items of central interest.

Three different exemplar images were collected for each location, place type, and situation, making a total 432 different images and 144 different labels in the stimulus set. In the location condition, the exemplars were 3 different views of the same location, whereas in the category and situation conditions the exemplars were 3 different scenes (e.g. 3 different bedrooms). Each image was presented once and each label 3 times during the course of the experiment. Identification task (location identification vs. category identification vs. situation identification) was crossed with image presentation time (30 vs. 80 ms) in a 3×2 design. Labels and images matched on 50% of the trials; labels and images on mismatch trials were different exemplars from the same stimulus category (location, place type, or situation).

Functional localizer scans were 8 m 12 s in length and divided into 16-s long picture epochs during which subjects viewed digitized color photographs of faces, common objects, scenes, and other stimuli presented at a rate of 1.25 pictures/s in a blocked design as described previously (Epstein and others 2005). Custom HRF scans were 8 m long. At 20-s intervals subjects were presented with a complex visual scene for 1 s and responded by pressing the button box with both thumbs in order to elicit a time-locked visuomotor response. The same parameters were used for functional localizer and custom HRF scans in all 4 experiments.

Experiment 2

Scan sessions consisted of 6 experimental scans of identical length to those in experiment 1 followed by 2 functional localizer scans and 1 custom HRF scan. Scans 1-3 included 72 picture-only trials and 72 null trials. The former had the same event structure as the label + picture trials presented in experiment 1 except that no verbal label preceded the image. Subjects were instructed to simply press a button to indicate whether or not they could successfully perceive each briefly presented picture; thus, the same nonspecific task was applied to all stimuli. Scans 4-6 included 72 label + picture trials (repeating the procedure of experiment 1) and 72 null trials. Picture-only trials were presented before label + picture trials to reduce the chance that subjects would apply different identification strategies to different image categories in picture-only trials. Images corresponding to half of the 48 Penn locations, place types, and situations were presented in the picture-only trials, whereas images and labels corresponding to the other half of the stimulus set were presented in label + picture trials. These assignments were randomized across subjects. Label presence (image-only vs. label + picture) was crossed with image set (locations vs. place types vs. situations) and image presentation time (30 vs. 80 ms) in a $2 \times 3 \times 2$ design.

Experiment 3

Scan sessions consisted of 8 experimental scans followed by 2 functional localizer scans and 1 custom HRF scan. Experimental scans were 6 m 48 s long and divided into 72 stimulus trials randomly interleaved with 24 null trials. A smaller number of null trials were used in this case to keep the experiment at a reasonable length given the larger number of

stimulus trials. The 72 stimulus trials in each scan included 36 label-only trials and 36 label + picture trials, which had the same event structure except that a scrambled-picture mask was presented in lieu of an image in label-only trials. Thus, response during label-only trials reflects task-related processing that occurs in the absence of a coherent visual scene. Label-only and label + picture trials were randomly interleaved within each scan to ensure that subjects could not predict whether or not an image would appear on any given trial; consequently, subjects had to initiate the appropriate identification task defined by the label on all trials. Subjects used a button box to report whether 1) a coherent image was presented that matched the label, 2) a coherent image was presented that did not match the label, and 3) only scrambled masks were presented. Image presence (label-only vs. label + picture) was crossed with identification task (location identification vs. category identification vs. situation identification) in a 2×3 design. Each of the 144 labels appeared twice during the experiment followed by an image and twice followed by a mask. Half of the images in label + picture trials were presented for 30 ms, whereas the other half were presented for 80 ms.

Experiment 4

Scan sessions consisted of 6 experimental scans followed by 2 functional localizer scans and 1 custom HRF scan. Experimental scans were 7 m 36 s long and divided into 48 label-only trials, 48 label + picture trials, and 24 null trials, all of which were randomly interleaved. Labels in scans 1-3 were names of specific Penn locations or place categories as in the previous 3 experiments. In contrast, labels in scans 4-6 were generic task cues ("Where is this?" for location identification, and "What is this?" for category identification). These generic labels were intended to prompt subjects to prepare to identify images in terms of either their location or category without providing specific information about what the images might be. For generic-label trials, subjects were instructed to report whether 1) a coherent image was presented that they could successfully interpret according to the instructions, 2) a coherent image was presented that they could not successfully interpret, and 3) only scrambled masks were presented. Label type (generic vs. specific) was crossed with image presence (label-only vs. label + picture) and identification task (location identification vs. category identification) in a $2 \times 2 \times 2$ design. Labels and images corresponding to half of the 48 Penn locations and place types were presented with specific labels in scans 1-3, whereas images corresponding to the other half were presented with generic labels in scans 4-6 (assignments randomized across subjects). Half of the images in label + picture trials were presented for 30 ms, whereas the other half were presented for 80 ms. Note that this experiment only examined location and category identification; to maximize the power available in these conditions, situation identification trials were not included.

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 of the scan, spatially normalized to the Montreal Neurological Institute (MNI) template, resampled into 3-mm isotropic voxels and spatially smoothed with an 8-mm full-width half maximum (FWHM) 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. Each stimulus condition was modeled as an impulse response function (experimental scans) or a boxcar function (functional localizer scans) convolved with a subject-specific HRF. In addition, estimated event-related time courses (e.g. Fig. 3) were calculated for display purposes by modeling each time point as a finite impulse response in a separate general linear model. Both region of interest (ROI) and whole brain analyses were performed.

For ROI analyses, data from the functional localizer scans were used to identify subject-specific regions responding more strongly to scenes than to common objects in the posterior parahippocampal/collateral sulcus region (PPA) and RSC. In addition, regions responding more strongly to scenes than to common objects were identified in the transverse occipital sulcus (TOS), more strongly to faces than to objects in the lateral fusiform gyrus (fusiform face area [FFA]) and more strongly to objects than to scenes in the lateral occipital (LO) region. Thresholds were set for each region in a subject-by-subject manner so that the ROIs were consistent with those identified in previous studies. In particular, for each subject, we chose the highest threshold that was low enough so that the ROIs could be reliably identified. These subject-specific thresholds ranged from $t = 2.0$ to $t = 6.0$, with the most common value of $t = 3.5$. PPA, RSC, TOS, and LO were defined bilaterally, whereas the FFA was only defined in the right hemisphere reflecting the typical right-sided lateralization of this region. Mean sizes for the scene-selective ROIs were: left PPA $2.9 \pm 1.2 \text{ cm}^3$, right PPA $3.2 \pm 1.2 \text{ cm}^3$, left RSC $1.5 \pm 1.3 \text{ cm}^3$, right RSC $1.8 \pm 1.1 \text{ cm}^3$, left TOS $1.7 \pm 1.1 \text{ cm}^3$, right TOS $1.8 \pm 1.0 \text{ cm}^3$. The range extended from just a few voxels in some cases up to 6.2 cm^3 .

The time course of MR response during the main experimental scans was extracted from each ROI (averaging over all voxels) and entered into the general linear model in order to calculate parameter estimates (beta values) for each condition, which were used as the dependent variables in a second-level random effects analysis of variance (ANOVA). Hemisphere was included as a factor in the omnibus ANOVAs for scene-responsive regions. Patterns of interest did not vary qualitatively between hemisphere although response differences between recognition tasks were larger in the left than the right PPA in experiment 2 ($F_{2,14} = 4.6$, $P < 0.05$) and experiment 3 ($F_{2,12} = 4.8$, $P < 0.05$) and the advantage for label + picture over label-only trials was larger in the right than in the left RSC in experiment 3 ($F_{1,5} = 7.3$, $P < 0.05$). Data were averaged across hemisphere for targeted analyses (i.e. ANOVAs and t -tests that specifically examined the location vs. category and category vs. situation effects).

For whole-brain analyses, subject-specific beta-maps were calculated for contrasts of interest and then smoothed to 12-mm FWHM to facilitate between-subject averaging before entry into a random effects analysis. A permutation analysis was used to control the family-wise error rate at $P < 0.05$ (2-tailed) corrected for multiple comparisons across voxels. In this analysis, an estimate of the distribution of brain-wise maximum t -values under the null hypothesis was calculated by randomly assigning positive or negative signs to the contrast of interest for each subject (Nichols and Holmes 2002). Data were then extracted from clusters of 10 or more significant voxels and analyzed as described above.

Subject-specific HRFs were obtained by using a Fourier basis set to identify voxels whose response was reliably affected by visuomotor stimulation in the custom HRF scan. The time-locked response to the stimulus impulse was then averaged across all responsive voxels.

Results

Experiment 1

fMRI responses in the PPA and RSC were strongly affected by the identification task (PPA: $F_{2,8} = 22.0$, $P < 0.002$; RSC: $F_{2,8} =$

22.0 , $P < 0.002$). Specifically, response to location identification trials was significantly higher than response to category identification and situation identification trials in both PPA and RSC (t values > 4.6 , P values < 0.01), but responses to category and situation identification trials did not differ (t values < 1.7 , P values > 0.15 , not significant [NS]) (Fig. 3a). The preferential response to location identification trials was quite striking, particularly in the RSC, where the response to location identification trials was 4 times the response to category identification trials. Both regions responded more strongly to 80-ms image presentations than to 30-ms image presentations (PPA: $F_{1,4} = 11.0$, $P < 0.05$; RSC: $F_{1,4} = 22.0$, $P < 0.01$); however, there was no interaction between presentation time and trial type (PPA: $F_{1,4} = 1.5$, $P = 0.27$, NS; RSC: $F_{1,4} = 1.9$, $P = 0.20$, NS).

These results suggest that the PPA and RSC are strongly sensitive to recognition level when identifying places. However, interpretation of these data is complicated by the fact that different image sets were used in the location and category identification conditions. As such, the results could potentially be due to physical differences between the stimulus sets, or to the greater familiarity of the places depicted in the location identification images (although see Epstein and others 1999). Experiment 2 was designed to control for these possibilities.

Experiment 2

In this experiment, subjects were presented with the photographs used in experiment 1 but without preceding labels (picture-only trials). This allowed us to measure the response to the images alone in a context where subjects were not required to perform different recognition tasks. Later in the same scan session, subjects viewed photographs with preceding labels and reported whether the pictures matched the labels, repeating the procedure of experiment 1 (label + picture trials).

The results (Fig. 3b) show a clear pattern: the significant main effects of image set (PPA: $F_{2,14} = 13.4$, $P = 0.001$; RSC: $F_{2,12} = 54.3$, $P < 0.001$) were modulated by significant interactions between label presence and image set (PPA: $F_{2,14} = 6.6$, $P = 0.01$; RSC: $F_{2,12} = 27.5$, $P < 0.001$). Specifically, response to location images was higher than response to category images when the images were preceded by labels that prompted subjects to perform different identification tasks (PPA: $t(7) = 5.2$, $P < 0.001$; RSC: $t(7) = 6.1$, $P < 0.001$) but not when they were presented alone (t values < 1 , NS). These results replicate the main findings of experiment 1 and demonstrate that the location identification versus category identification response differences in PPA and RSC cannot be attributed to the use of different image sets in the 2 conditions. Rather, these differences appear to be driven by processes induced by the verbal labels. We test this hypothesis further in experiment 3.

A secondary effect found in experiment 2 (but not experiment 1) was greater response during category identification than during situation identification in the PPA and RSC (Fig. 3b). This difference was found for both label + picture (PPA: $t(7) = 2.6$, $P < 0.05$; RSC: $t(7) = 3.2$, $P < 0.02$) and picture-only trials (PPA: $t(7) = 2.8$, $P < 0.05$; RSC: $t(7) = 3.7$, $P < 0.01$). Indeed, the category versus situation difference was not modulated by label presence in the PPA ($F < 1$, NS) or the RSC ($F_{1,7} = 3.2$, $P = 0.12$, NS). This contrasts with the location versus category effect, which was significantly larger for label + picture than for picture-only trials in both regions (PPA: $F_{1,6} = 7.5$, $P < 0.05$; RSC: $F_{1,7} = 28.9$, $P = 0.001$). This pattern suggests that the category versus situation difference is attributable to qualitative

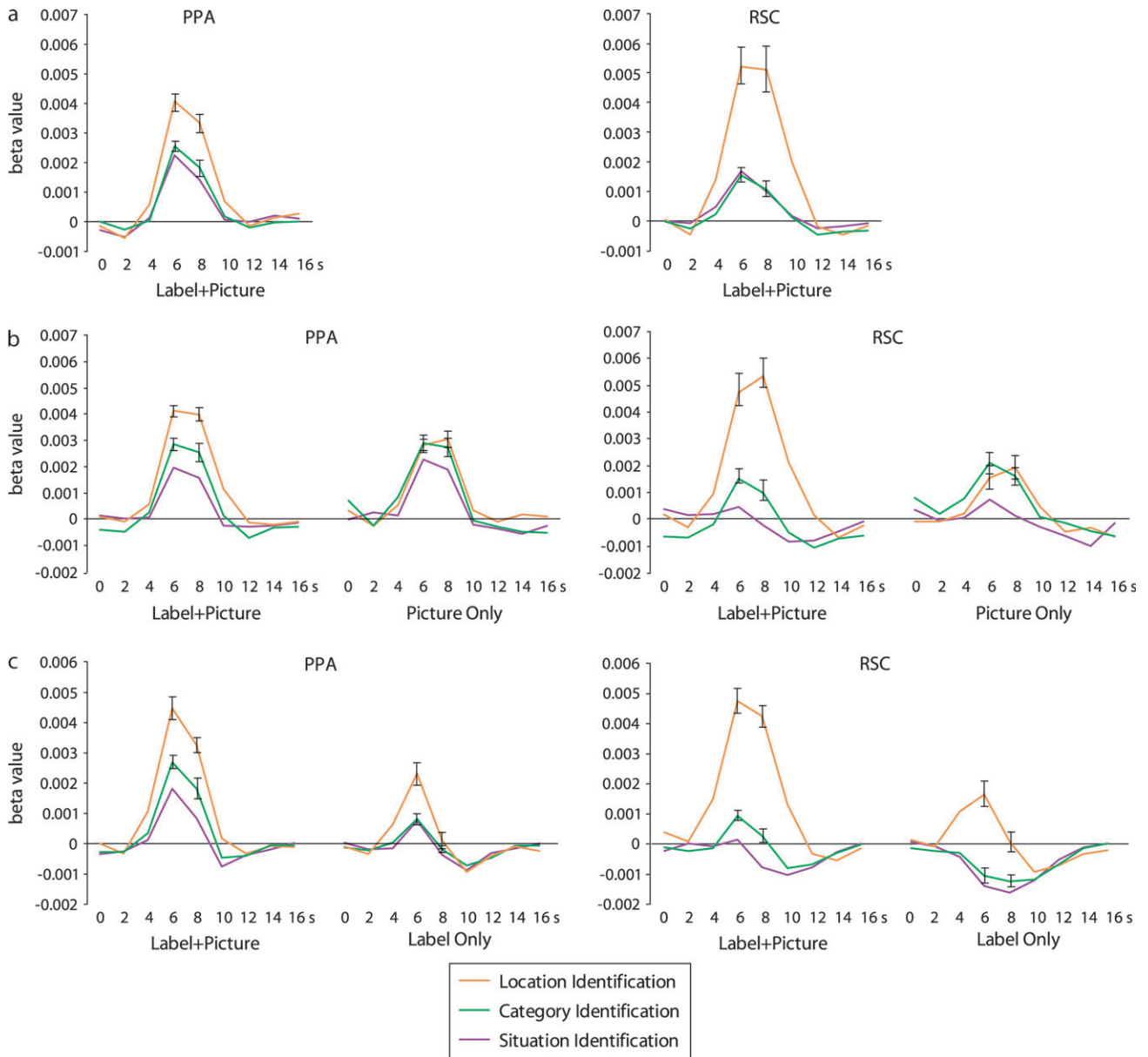


Figure 3. Event-related response in the PPA and RSC in experiments 1–3. Time courses were estimated using a finite impulse response model; statistics reported in the paper reflect effect sizes estimated using a standard HRF, which gave the same pattern of results. Error markers on the location identification curve reflect ± 1 standard error of the mean (SEM) for the location versus category identification difference at the 6-s and 8-s time points, whereas error bars on the category identification curve reflect ± 1 SEM for the category versus situation identification difference. Units on the y -axis are beta values. (a) Experiment 1: Response during location identification trials was significantly higher than response during category and situation identification trials in both regions. (b) Experiment 2: The location advantage was observed on label + picture trials but not on picture-only trials in which images were presented alone without preceding labels. Note that there was also a significant advantage for category versus situation identification trials, which did not depend on label presence and may reflect preferential response to the category identification images. (c) Experiment 3: The location versus category identification difference was observed on both label + picture and label-only trials, demonstrating that verbal cues can drive this effect even in the absence of a coherent visual scene. In contrast, the category versus situation identification difference is only found on label + picture trials.

differences between the 2 image sets. In particular, the images in the category and location identification conditions generally contained unobstructed views of background architectural features, whereas the images in the situation condition generally contained highly salient interacting people in the foreground. Thus, these results indicate that PPA and RSC respond more strongly to scenes in which the enduring features that define the spatial context of the scene are salient; however, this image-based effect appears to be smaller and less reliable than the effect of recognition level.

Experiment 3

The preceding results indicate that response differences between the location and category identification conditions are driven by processes induced by the verbal labels. There are at least 2 different ways in which the labels might have affected recognition. First, they might have conditioned the recognition system prior to the appearance of the image, by inducing attentional or retrieval processes such as orientation to particular parts of the visual field (Chun and Jiang 1998), or formation of a mental image of the named location or category (Ishai and

others 2000; O'Craven and Kanwisher 2000; Kosslyn and others 2001). Second, they might have conditioned the recognition system during and after the appearance of the image, either by affecting visual processing directly or by inducing retrieval of different semantic memory codes corresponding to either specific familiar locations or general place categories. Note that these possibilities are not mutually exclusive: the use of labels of different types might have affected neural activity both before and after the presentation of the images.

In order to clarify the nature of the label-induced processes, experiment 3 compared response during label + picture trials with response during label-only trials in which no picture appeared. We hypothesized that label-cued attentional or imagery processes should engage on both label + picture and label-only trials, whereas label-cued differences in perceptual processing or mnemonic retrieval should only be found on label + picture trials. Consequently, there were 2 questions of interest: 1) Would the location advantage previously observed on label + picture trials also be found on label-only trials (indicating label-cued attentional or imagery processes)? 2) If the location advantage was found for both trial types would it be significantly larger on label + picture than on label-only trials (indicating the additional involvement of postimage location recognition processes)?

The results are plotted in Figure 3(c). ANOVA found significant main effects of identification task (PPA: $F_{2,6} = 19.6$, $P < 0.001$; RSC: $F_{2,6} = 195.2$, $P < 0.001$) and image presence (PPA: $F_{1,6} = 56.9$, $P < 0.001$; RSC: $F_{1,6} = 80.7$, $P < 0.001$). Critically, planned *t*-tests revealed greater response during location identification than during category identification for both label + picture (PPA: $t(6) = 4.9$, $P < 0.005$; RSC: $t(6) = 11.0$, $P < 0.001$) and label-only trials (PPA: $t(6) = 3.3$, $P < 0.02$; RSC: $t(6) = 6.1$, $P < 0.001$), demonstrating that activity in both regions can be driven by the labels even in the absence of an image. This location advantage was significantly larger for label + picture trials than for label-only trials in RSC ($F_{1,6} = 9.9$, $P < 0.05$) but not in the PPA ($F_{1,6} = 3.5$, $P = 0.11$, NS). These apparent regional differences were confirmed by an additional ANOVA, which found a significant triple interaction of ROI (PPA vs. RSC), label presence (label + picture vs. label-only), and task (location vs. category) ($F_{1,6} = 8.7$, $P < 0.05$). Thus, most of the location advantage in the PPA seems to reflect label-triggered attentional or imagery effects that engage on location trials irrespective of whether or not a coherent image subsequently appears. Although these effects are also found in the RSC, at least some of the location advantage in this region may reflect postimage processing that only occurs on label + picture trials.

As in experiment 2, we observed greater response to category identification than to situation identification in label + picture trials in the PPA ($t(6) = 4.4$, $P < 0.01$) and RSC ($t(6) = 5.6$, $P = 0.001$). This difference was not observed on label-only trials (PPA: $t < 1$, NS; RSC: $t(6) = 1.7$, $P = 0.14$, NS), consistent with the idea that this effect is driven by physical differences between the images used in these 2 conditions.

Experiment 4

To confirm the hypothesis that the verbal labels induce different types of processes in the PPA and RSC, we ran a fourth experiment in which 2 kinds of labels were used: labels cueing recognition of a specific location or type of place ("38th St. Bridge," "kitchen") or labels that generically cued these 2

recognition tasks ("Where is this?" "What is this?"). We reasoned that the generic labels—although adequate as task cues—would not induce preimage retrieval of information about specific locations. Thus, we predicted that the location advantage in the PPA would be largely eliminated by the use of generic labels, whereas the location advantage in RSC would be reduced but not eliminated.

These predictions were borne out by the results (Fig. 4). ANOVA revealed a location advantage in both regions (PPA: $F_{1,7} = 26.5$, $P < 0.002$; RSC: $F_{1,7} = 35.0$, $P < 0.001$); which was larger for specific labels than for generic labels (PPA: $F_{1,7} = 10.9$, $P < 0.02$; RSC: $F_{1,7} = 15.0$, $P < 0.01$). Additional significant main effects included greater response to label + picture trials than to label-only trials (PPA: $F_{1,7} = 72.0$, $P < 0.001$; RSC: $F_{1,7} = 57.4$, $P < 0.001$) and greater response for specific labels than for generic labels (PPA: $F_{1,7} = 9.2$, $P < 0.02$; RSC: $F_{1,7} = 14.6$, $P < 0.01$). Planned *t*-tests found that generic labels induced a location versus category response difference in the RSC, both when these labels were followed by a coherent image on label + picture trials ($t(7) = 2.9$, $P < 0.05$) and when they were followed by a mask on label-only trials ($t(7) = 3.7$, $P < 0.01$). In contrast, no location identification advantage was observed in the PPA when generic labels were used, either on label + picture ($t(7) = 1.6$, $P = 0.16$, NS) or on label-only trials ($t < 1.1$, NS). These regional differences were confirmed by an additional ANOVA performed on generic-label trials with ROI (PPA vs. RSC), image-presence (label + picture vs. label-only), and task (location vs. category identification) as factors. For these trials, the location versus category identification difference was significantly larger in the RSC than in the PPA ($F_{1,7} = 19.5$, $P < 0.01$).

In sum, the results demonstrate that generic verbal labels are efficacious in the RSC but not in the PPA, supporting the contention that some of the location advantage in RSC may reflect retrieval of information about the larger spatial context within which the depicted scene is to be placed. In contrast, the location advantage in the PPA may reflect attentional or imagery processes that are cued by labels describing specific locations but not by generic verbal task cues.

Other Cortical Regions

Although this study primarily focused on the PPA and RSC, we also performed a whole-brain analysis in order to answer 3 questions that could not be addressed with ROI analyses. First, is the recognition level effect (i.e. greater response during location identification than during category identification) specific to the PPA and RSC, or can the same effect be observed in many other brain regions? Second, are there any regions that exhibit the opposite pattern, of greater response during category identification than during location identification? Third, are there any regions that respond more strongly to situational scenes than to the spatial scenes observed during location and category identification, complementing the greater response of the PPA/RSC to spatial scenes? To maximize the power available for these whole-brain analyses, data from the label + picture conditions in experiments 1–3 were combined. Data from experiment 4 were not included because there was no situation identification condition in this experiment.

Regions that responded differentially to location versus category identification at the corrected significance level are listed in Table 1. The strongest responses to location identification trials were found in parahippocampal and retrosplenial

cortices, in almost the exact same loci as those defined from the functional localizer data. Preferential response during location identification was also observed in the TOS and right primary visual cortex. The TOS has been previously reported to respond preferentially to scenes (Grill-Spector 2003; Hasson and others 2003; Epstein and others 2005). A separate ROI analysis performed on scene-responsive voxels in the TOS found a similar pattern to the PPA in experiments 1-3 (data not shown). The V1 activation appears to be focused on part of V1 that represents the periphery of the visual field, suggesting that subjects may attend more strongly to the periphery of the visual scene during location identification (Brefczynski and DeYoe 1999; Somers and others 1999; Levy and others 2001). In general, these results support the claim that the PPA and RSC are the regions of the brain most involved in location identification, although a similar pattern can be observed in other regions.

Three regions responded more strongly during category identification than during location identification (Table 1 and

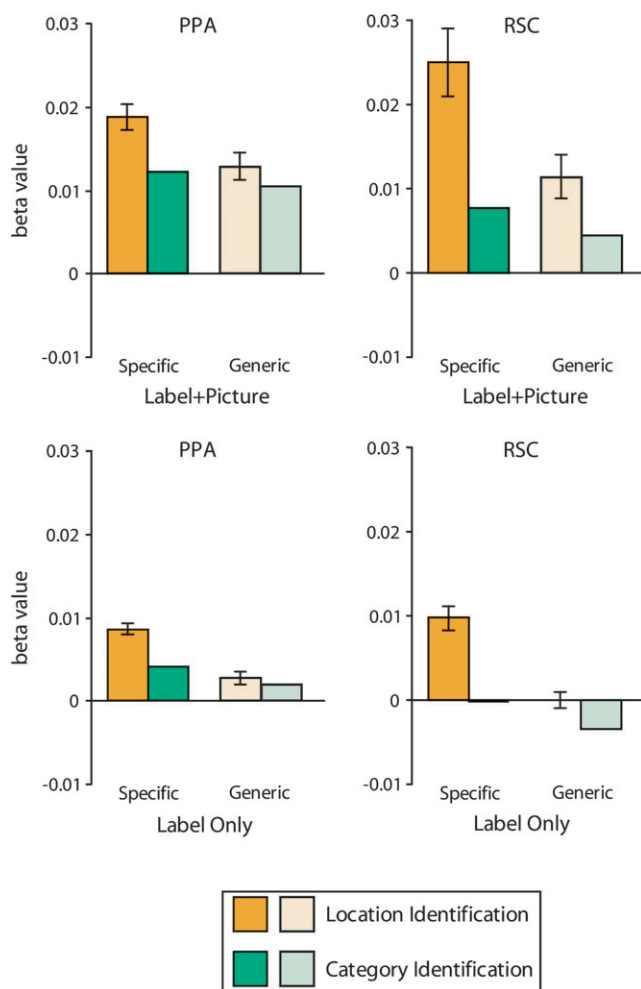


Figure 4. Response in the PPA and RSC in experiment 4. When specific labels were used (solid bars) both the PPA and RSC responded more strongly during location identification than during category identification for both label + picture trials (top panels) and label-only trials (bottom panels), replicating the results of experiment 3. This preferential response during location identification was significantly reduced when generic labels ("Where is this?," "What is this?") were used (lighter bars), although significant differences were still observed in the RSC. Error markers on the location arms condition reflect ± 1 standard error of the mean for the location versus category identification difference; units on the y-axis are beta values.

Fig. 5): right superior temporal sulcus (STS), left supermarginal gyrus (SMG), and the left cingulate sulcus (CiS). The fact that some regions respond more strongly during category identification suggests the existence of 2 partially dissociable cortical systems involved in location identification and place category identification, rather than a single system that is more engaged when scenes must be identified at the specific rather than the basic level. However, the involvement of the right STS, left SMG, and CiS in category identification may be incidental to their main function, as the response in these regions during category identification was no stronger than their response to the intertrial baseline. Interestingly, all of these regions responded at least as strongly during situation identification trials as they did during category identification trials, even though they were defined without reference to their situation identification response. Indeed, response in the STS to situation identification was *stronger* than response during category identification ($t(19) = 3.6, P < 0.002$). We consider the implication of these results further in the discussion.

The observation that primary visual cortex responds more strongly during location identification than during category identification raises the concern that some of the preferential response to location identification in the PPA and RSC might simply reflect the greater difficulty of this task, which might require the recruitment of additional attentional resources. To ensure that any such attentional effects are not general across all visual areas, we performed an ROI analysis on 2 functionally defined areas that are not expected to be involved in place recognition: the FFA and LO. Previous experiments have demonstrated that the response to scenes in these regions is about 50% as strong as the response to their preferred stimulus (Downing and others 2006; Grill-Spector 2003). As expected, neither of these regions responded differentially to location versus category identification (both t values < 1 , NS; see Fig. 6). Thus, effects of recognition level were not found in all high-level visual areas but were largely restricted to the PPA and RSC. Interestingly, greater response during situation identification than category identification was observed in both the FFA ($t(19) = 5.2, P < 0.0001$) and LO ($t(19) = 2.9, P < 0.02$), probably reflecting preferential response to the faces and bodies in this condition.

A whole-brain analysis of the scene type effect (spatial vs. social scenes) found only one region that responded differentially to the category identification and situation identification conditions at the corrected significance level. This was the left STS, which responded more strongly during situation identification (Fig. 5). However, when the significance level was relaxed, a more complete pattern emerged, including greater response during situation identification in the left and right fusiform gyrus (corresponding to the FFA) and right STS, and greater response during category identification in the PPA. These results are consistent with earlier work that implicates the fusiform gyrus in face/body processing (Puce and others 1996; Kanwisher and others 1997; Peelen and Downing 2005) and STS in the analysis complex social stimuli (Puce and others 1998; Haxby and others 2000; Pelphrey and others 2004; Saxe and others 2004).

Behavioral Data

Behavioral data for all experiments are reported in Tables 2 and 3. Subjects took longer to respond and made more errors on location identification trials than on category or situation

identification trials in all 4 experiments. However, 4 lines of evidence suggest that these behavioral differences are unlikely to explain the fMRI effects. First, PPA and RSC response patterns during label + picture trials did not change when response time was added as a covariate in experiments 1 and 2 (data not shown). Second, an advantage for location identification was observed in label-only trials in experiment 3, even though no response time differences were observed. Third, response time differences cannot account for the several regions that re-

Table 1
Regions showing significant effects ($P < 0.05$, 2-tailed, corrected for multiple comparison) of recognition level (location vs. category identification) or scene type (category vs. situation identification) in a whole-brain analysis

Region and contrast	MNI coordinates			z-score
	x	y	z	
Location ID > category ID				
L posterior parahippocampal (PPA)	-19	-37	-8	5.71
R posterior parahippocampal (PPA)	20	-36	-6	5.88
L retrosplenial	-10	-59	8	6.11
R retrosplenial	13	-54	9	6.52
L TOS	-33	-79	31	5.10
R TOS	32	-75	34	4.89
R primary visual cortex	13	-92	4	6.52
Category ID > location ID				
L CIS	-10	-21	51	4.91
R STS	56	-63	7	4.55
L SMG	-60	-28	40	4.48
Situation ID > category ID				
L STS	-50	-58	15	5.14

sponded more strongly during situation and category identification than during location identification in the whole brain analysis. Finally, we performed an additional analysis on the label + picture trials in which correct trials, incorrect trials, and trials with no response were modeled separately. The results for the correct trials are plotted in Supplementary Figure 1, along with the original results (which included all trials). The results do not change appreciably when only correct trials are considered, demonstrating that the response patterns cannot be attributed to accuracy differences between the conditions. Note, however, that it remains possible that location identification may have greater or different additional requirements than the other tasks, leading to both higher fMRI response and lower accuracy in this condition.

Discussion

The studies reported in this paper addressed 2 questions. First, are the neural systems involved in place recognition sensitive to the level (basic category vs. specific exemplar) at which places are identified? Second, are the neural systems involved in place recognition also engaged during situation recognition? The studies focused on the PPA and RSC, which previous neuroimaging and neuropsychological work have identified as being critical for scene recognition (Aguirre and D'Esposito 1999; Mendez and Cherrier 2003; Epstein 2005). We found that these regions were surprisingly sensitive to recognition level when identifying places, responding more strongly during location identification than during place category identification in all 4 experiments. Furthermore, these regions were sensitive to the

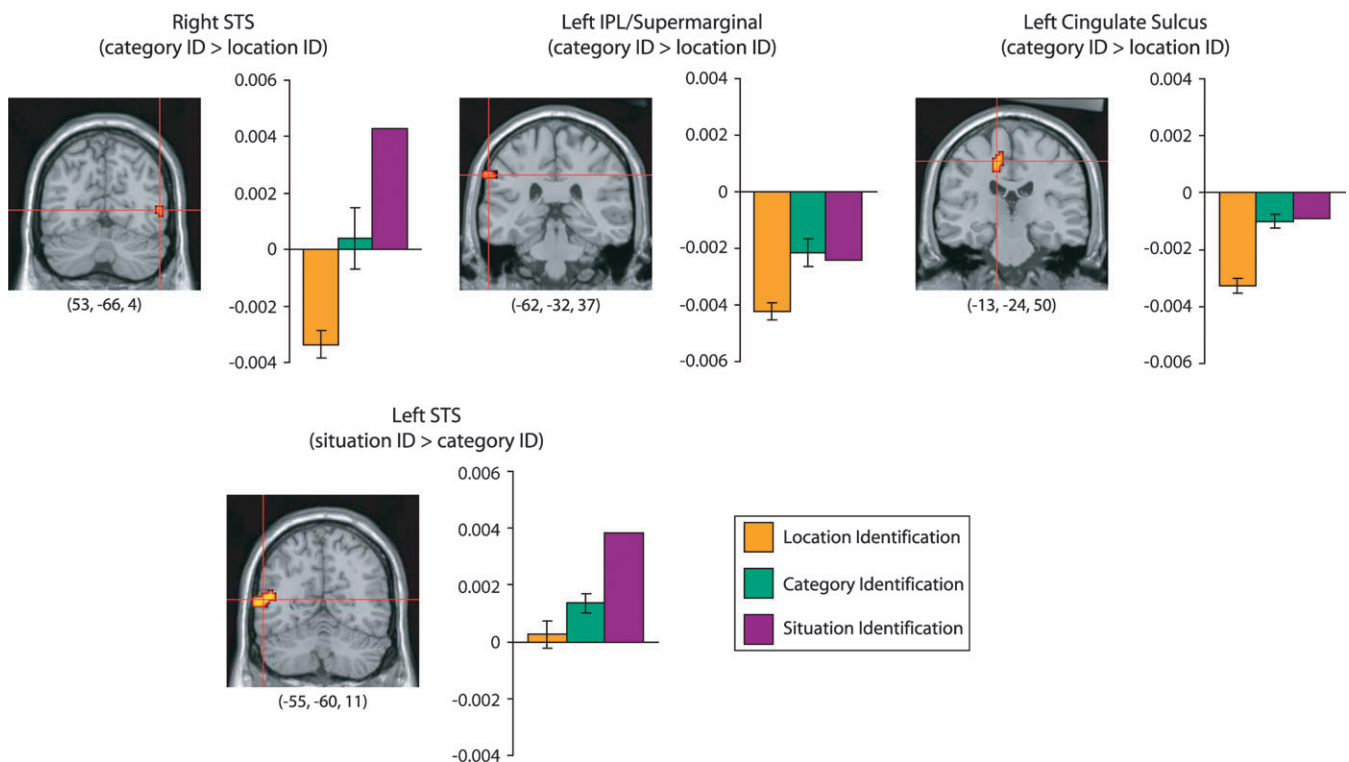


Figure 5. Partial results of the whole-brain group analysis (see also Table 1). Regions responding more strongly ($P < 0.05$, 2-tailed, corrected) during category identification trials than to location identification trials (top) or more strongly during situation identification than during category identification trials (bottom) are plotted. Data are combined for label + picture trials across experiments 1–3 and overlaid on a partially inflated template brain in MNI space; right hemisphere is on the right. Bar charts show the mean response to all 3 conditions in each region. Error markers on the location identification bars reflect ± 1 standard error of the mean (SEM) for the location versus category identification difference, whereas error markers on the category identification bars reflect ± 1 SEM for the category versus situation identification difference; units on the y-axis are beta values.

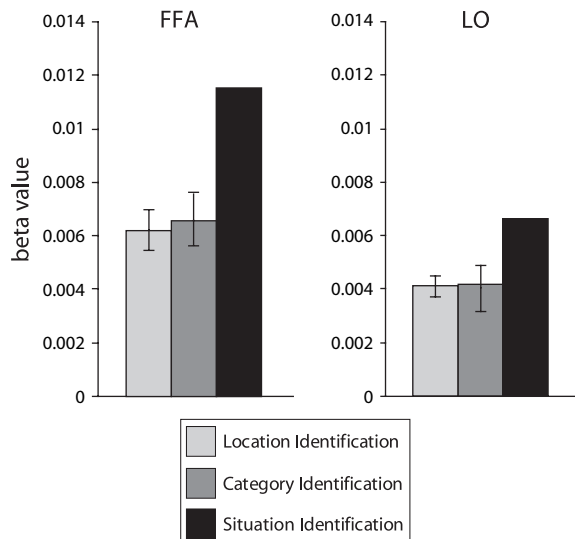


Figure 6. Response in the FFA and LO. Data are combined for label + picture trials across experiments 1–3. Neither region responded differentially to location versus category identification trials, demonstrating that not all high-level visual regions exhibit the location advantage observed in the PPA and RSC. Both regions responded more strongly to situation identification trials, likely reflecting preferential response to the bodies and faces shown in this condition. Error markers are the same as in Figure 4.

Table 2
Mean response times and standard errors of the mean for each experiment

Exp. 1	Location ID	Category ID	Situation ID	Mean		
Label + picture	921 ± 68	849 ± 55	804 ± 42	858 ± 54		
Exp. 2	Location ID	Category ID	Situation ID	Mean		
Label + picture	1045 ± 46	922 ± 61	867 ± 53	943 ± 51		
Picture-only	981 ± 67	976 ± 59	929 ± 53	962 ± 58		
Mean	1013 ± 55	949 ± 58	898 ± 52	953 ± 54		
Exp. 3	Location ID	Category ID	Situation ID	Mean		
Label + picture	1036 ± 70	1022 ± 67	935 ± 60	997 ± 64		
Label-only	966 ± 97	969 ± 91	966 ± 92	967 ± 92		
Mean	1001 ± 82	996 ± 77	951 ± 73	982 ± 77		
Exp. 4	Generic label			Specific label		
	Location ID	Category ID	Mean	Location ID	Category ID	Mean
Label + picture	1109 ± 50	1087 ± 44	1098 ± 45	1192 ± 23	1066 ± 29	1127 ± 23
Label-only	906 ± 49	878 ± 41	892 ± 45	1074 ± 46	1051 ± 51	1062 ± 48
Mean	1008 ± 34	983 ± 33	994 ± 33	1133 ± 27	1059 ± 27	1094 ± 26

type of scene being recognized, responding more strongly when places were identified than when situations were identified (both at the categorical level) in 2 out of the 3 experiments in which both of these conditions were included. These results provide new insights into the neural mechanisms underlying scene recognition by demonstrating differential involvement of the PPA and RSC in different scene identification tasks.

In the basic paradigm, repeated in all 4 experiments, subjects were presented with a label naming a place category, location, or situation, and reported whether the label matched a subsequently presented visual scene. Results from several variants of this paradigm suggest that the recognition level effect (i.e. greater response during location identification than during place category identification) was driven by processes induced by the verbal labels. The location advantage was not observed when images were presented without preceding labels in experiment 2, but was observed when labels were presented

without succeeding images in experiment 3. Thus, this effect cannot be attributed to visual differences between the image sets or to greater response to images of familiar places than to images of unfamiliar places because it did not occur when subjects simply viewed the images without performing different recognition tasks, but did occur when subjects prepared to perform different recognition tasks but did not view the images. The location advantage was also observed in RSC (but not PPA) in experiment 4 when generic labels were used. These results fit with previous studies that have suggested that neural responses in some areas of high-level visual cortex may be as strongly driven by the type of information retrieved in response to a stimulus as they are by the visual qualities of the stimulus itself (Martin and others 1996; Cox and others 2004). In the current experiment, the labels might have been particularly efficacious relative to the images because of their much longer presentation times (1 s for the labels vs. 30/80 ms. for the images).

In contrast, the scene type effect (i.e. greater response during place category identification than during situation identification) appears to be driven by physical differences between the image sets, rather than by processes induced by the verbal labels. This effect was observed on both label + picture and picture-only trials in experiment 2 but was not observed on label-only trials in experiment 3. A key difference between the images used in the situation identification and place category conditions is the fact that the former images contained highly salient foreground elements (usually people), whereas the latter images generally displayed unobstructed views of fixed background elements such as walls, furniture, and buildings. Given the brief presentation times used in this experiment, subjects may have attended more strongly to the foreground elements of the situation images, ignoring the background elements that define the scene as a place. The fact that the foreground elements were generally people rather than inanimate objects may have additionally reduced the response of the PPA, as previous studies have indicated that this region responds less strongly to biological stimuli such as faces (Epstein and Kanwisher 1998), animals (Chao and others 1999), and bodies (Downing and others 2005) than to inanimate stimuli. In any case, these results indicate that the PPA and RSC are unlikely to support general visual integration mechanisms that mediate recognition of both spatial and social scenes.

Returning to the recognition level effect, we observed certain notable differences in the ways this effect is manifested in the PPA and RSC that point to different functional roles for these 2 regions during place recognition. First, the location identification versus category identification difference was larger in the RSC than in the PPA (400% vs. 70% increase in label + picture trials), suggesting that RSC was more affected by the recognition level manipulation than the PPA. Second, in experiment 3, the location versus category difference was larger for label + picture than for label-only trials in RSC but not in the PPA, suggesting that the RSC is more likely than the PPA to support postrecognition semantic retrieval or orientational processes (which would only engage in trials where both a label and an image appear). Third, in experiment 4, only the RSC exhibited significant location versus category differences when these tasks were cued with generic rather than specific labels. These results suggest that the PPA and RSC may play distinct but complementary roles in place recognition.

In particular, we hypothesize that the PPA may support representations of the visuospatial structure of individual

Table 3

Mean accuracy and standard errors of the mean for each experiment; see Results section for task descriptions

Exp. 1		Location ID	Category ID	Situation ID	Mean	
Label + picture	Correct	0.73 ± 0.03	0.80 ± 0.02	0.83 ± 0.03	0.79 ± 0.03	
	Incorrect	0.22 ± 0.03	0.16 ± 0.01	0.13 ± 0.01	0.17 ± 0.01	
	No response	0.05 ± 0.02	0.04 ± 0.01	0.04 ± 0.02	0.04 ± 0.02	
Exp. 2		Location ID	Category ID	Situation ID	Mean	
Label + picture	Correct	0.64 ± 0.05	0.70 ± 0.08	0.76 ± 0.08	0.70 ± 0.07	
	Incorrect	0.19 ± 0.02	0.14 ± 0.02	0.12 ± 0.02	0.15 ± 0.01	
	No response	0.16 ± 0.07	0.16 ± 0.08	0.12 ± 0.06	0.15 ± 0.07	
Picture-only	Recognized	0.46 ± 0.06	0.44 ± 0.06	0.63 ± 0.05	0.51 ± 0.05	
	Not recognized	0.47 ± 0.05	0.49 ± 0.05	0.31 ± 0.04	0.42 ± 0.05	
	No response	0.08 ± 0.04	0.07 ± 0.04	0.07 ± 0.04	0.07 ± 0.04	
Exp. 3		Location ID	Category ID	Situation ID	Mean	
Label + picture	Correct	0.62 ± 0.04	0.75 ± 0.05	0.80 ± 0.04	0.72 ± 0.04	
	Incorrect	0.23 ± 0.02	0.14 ± 0.02	0.13 ± 0.02	0.17 ± 0.01	
	No response	0.10 ± 0.03	0.05 ± 0.02	0.04 ± 0.02	0.06 ± 0.02	
	Not detected	0.05 ± 0.02	0.06 ± 0.03	0.03 ± 0.01	0.05 ± 0.02	
Label-only	Correct rejection	0.52 ± 0.14	0.53 ± 0.14	0.53 ± 0.14	0.52 ± 0.14	
	False alarm	0.41 ± 0.13	0.40 ± 0.13	0.40 ± 0.13	0.41 ± 0.13	
	No response	0.07 ± 0.02	0.08 ± 0.03	0.07 ± 0.02	0.07 ± 0.02	
Exp. 4		Generic label		Specific label		Mean
		Location ID	Category ID	Location ID	Category ID	
Label + picture	Correct/recognized	0.54 ± 0.05	0.63 ± 0.06	0.65 ± 0.03	0.76 ± 0.04	0.65 ± 0.04
	Incorrect/not recognized	0.35 ± 0.05	0.26 ± 0.05	0.19 ± 0.02	0.11 ± 0.02	0.23 ± 0.02
	No response	0.09 ± 0.02	0.09 ± 0.02	0.13 ± 0.03	0.09 ± 0.02	0.10 ± 0.02
	Not detected	0.02 ± 0.02	0.03 ± 0.03	0.04 ± 0.02	0.05 ± 0.03	0.03 ± 0.02
Label-only	Correct rejection	0.76 ± 0.06	0.76 ± 0.06	0.60 ± 0.09	0.66 ± 0.09	0.69 ± 0.07
	False alarm	0.18 ± 0.06	0.16 ± 0.06	0.28 ± 0.07	0.27 ± 0.08	0.22 ± 0.06
	No response	0.06 ± 0.02	0.08 ± 0.02	0.12 ± 0.03	0.07 ± 0.02	0.08 ± 0.02

scenes (Epstein and Kanwisher 1998) that can be activated in response to a variety of visual and verbal cues (Maguire and others 1997; O'Craven and Kanwisher 2000; Kohler and others 2002; Janzen and van Turenout 2004), whereas the RSC may support more expansive representations that allow the currently-visible scene to be placed within a spatial framework that extends beyond the immediate horizon (McNamara and others 2003; Wolbers and Buchel 2005). Under this account, the PPA responds strongly whenever the fixed background elements that define the local spatial layout of a scene are viewed or imagined, even if the location of the scene is unknown (Epstein and others 1999). The additional boost of response observed in the PPA during location identification in the current study would reflect label-induced activation of specific geometric scene representations that occurs prior to the appearance of the image. As such, the magnitude of this location advantage is independent of whether or not a coherent image subsequently appears in the trial (as observed in Exp. 3). Furthermore, labels describing place categories (Expts 1-4) and generic task labels (Exp. 4) should not affect PPA response because neither of these labels induce retrieval of specific visuospatial scene representations. In RSC, this account predicts that response should be relatively weak when spatial scenes are merely viewed but much stronger when information about the larger environment surrounding the scenes can be retrieved. Location labels are hypothesized to have 2 effects in RSC: first, they act as item cues, inducing retrieval of information about the spatial setting of the named location; second, they act as task cues, inducing subjects to situate the subsequently appearing scene within the larger campus environment. The first effect accounts for the location advantage on label-only trials (Exp. 3), whereas the second effect accounts for the fact that the advantage is even larger on label + picture trials than on label-only trials (Exp.

3), and also for the location advantage observed on generic-label trials (Exp. 4).

This interpretation is consistent with several strands of research. First, neuropsychological studies have indicated that patients with parahippocampal damage have difficulty identifying individual scenes and landmarks (Habib and Sirigu 1987; Epstein and others 2001; Mendez and Chierri 2003), whereas patients with retrosplenial damage can identify scenes by name but cannot use them to orient in the wider world (Takahashi and others 1997; Aguirre and D'Esposito 1999; Katayama and others 1999; Maguire 2001). Second, neuroimaging studies have shown that RSC is particularly active during retrieval of information about the spatial structure of large-scale environments that extend across several distinct locations (Ino and others 2002; Wolbers and Buchel 2005), suggesting that this region represents stable aspects of the cognitive map in humans (Teng and Squire 1999; Rosenbaum and others 2000). Third, it has been established that PPA and RSC activity can be elicited by nonscene cues, including both verbal cues as in the current experiment (O'Craven and Kanwisher 2000) and in the case of the PPA landmark objects (Janzen and van Turenout 2004). Finally, computational models of navigation have demonstrated the utility of maintaining distinct representations of the local scene and of the larger environment, which can then be linked together by reference to common geometric features (Kuipers 2000; Kuipers and others 2004). The human brain might implement such a scheme by representing the local scene in the PPA and the global map in RSC.

Taken as a whole, our results suggest that the PPA and RSC are part of a neural system for identifying specific locations that is substantially less engaged by other scene recognition tasks. In particular, we hypothesize that these regions encode spatial information that is more useful for discriminating between

specific locations that have idiosyncratic geometric structures than for discriminating between basic-level place categories like “restaurant” or “library” whose geometric structures can vary widely across exemplars (Tversky and Hemenway 1983). However, we must bracket this proposal with 2 caveats. First, we cannot conclude from the results that the PPA and RSC are uninvolved in place category or situation identification, as these tasks might also tap PPA and RSC representations, but to a lesser degree than location identification. Second, we cannot exclude the possibility that some of the preferential PPA and RSC response to location identification trials might reflect retrieval of episodic or semantic information associated with specific familiar locations (Sugiura and others 2005). Future studies might clarify these issues in 2 ways. First, neuropsychological studies might determine whether parahippocampal and retrosplenial damage leads to impairment on both location identification and category identification tasks, or only on location identification tasks. Second, neuroimaging studies might determine whether the PPA and RSC respond differentially to retrieval of spatial versus nonspatial information about familiar places.

We also identified cortical regions that were sensitive to the recognition level and scene type manipulations but exhibited patterns opposite to those observed in the PPA and RSC. In particular, the right STS, left SMG, and left CiS responded more strongly during category identification than location identification, whereas the left STS responded more strongly during situation identification than during category identification. Examination of the response within these regions suggests that SMG and CiS do not distinguish between places and situations, responding equally to both when scenes are identified at the categorical rather than at the specific level. In contrast, STS responds maximally during situation identification, minimally during location identification, and at an intermediate level during category identification. Previous work has implicated STS in the processing of social information (Frith and Frith 2003; Saxe and others 2004; Kable and Chatterjee in press). Thus, one speculative possibility is that the brain answers the question “what kind of place is this?” in part by activating representations of the actions or events that might occur there. Alternatively, STS may serve as a multimodal integration area where visual features and sounds that are characteristic of different kinds of places, objects, and situations are represented (Amedi and others 2005; Beauchamp 2005). Further studies will be necessary to more fully delineate the network of regions involved in basic-level identification of social and spatial scenes. At present, we merely note that situation and category identification appear to recruit some regions that are significantly less engaged during location identification.

In sum, the current results advance our understanding of scene recognition by demonstrating that the PPA and RSC are part of a cortical system for identifying specific locations that is substantially less engaged by other scene recognition tasks, and by suggesting distinct roles for these regions in location identification. Results such as these may provide a first step toward the development of a neurally grounded theory of scene recognition that would complement and extend current theories of face and object recognition.

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Notes

We thank R. Caravella, K. Jablonski, D. Marshall, J. Mervis, A. Ponce De Leon, A. Shafir, and A. Soetanto for their assistance with these experiments. We also thank P. Downing, S. Thompson-Schill, M. Peelen, and 3 anonymous reviewers for useful comments on the manuscript. This research was supported by funds from the University of Pennsylvania School of Arts and Sciences and a grant from the Whitehall Foundation. *Conflict of Interest:* None declared.

Address correspondence to Russell Epstein, Department of Psychology, 3720 Walnut St, Philadelphia, PA 19104-6241, USA. Email: epstein@psych.upenn.edu.

References

- Aguirre GK, D'Esposito M. 1999. Topographical disorientation: a synthesis and taxonomy. *Brain* 122:1613–1628.
- Aguirre GK, Detre JA, Alsup DC, D'Esposito M. 1996. The parahippocampus subserves topographical learning in man. *Cereb Cortex* 6:823–829.
- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ. 2005. Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559–571.
- Bar M. 2004. Visual objects in context. *Nat Rev Neurosci* 5:617–629.
- Bar M, Aminoff E. 2003. Cortical analysis of visual context. *Neuron* 38:347–358.
- Beauchamp MS. 2005. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol* 15:145–153.
- Beauchamp MS, Lee KE, Haxby JV, Martin A. 2002. Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34:149–159.
- Biederman I. 1972. Perceiving real-world scenes. *Science* 177:77–80.
- Biederman I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94:115–147.
- Biederman I, Rabinowitz JC, Glass AL, Stacy EW Jr. 1974. On the information extracted from a glance at a scene. *J Exp Psychol* 103:597–600.
- Bohbot VD, Kalina M, Stepankova K, Spackova N, Petrides M, Nadel L. 1998. Spatial memory deficits in patients with lesions to the right hippocampus and to the right parahippocampal cortex. *Neuropsychologia* 36:1217–1238.
- Brefczynski JA, DeYoe EA. 1999. A physiological correlate of the ‘spotlight’ of visual attention. *Nat Neurosci* 2:370–374.
- Burgess N, Maguire EA, Spiers HJ, O'Keefe J. 2001. A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *Neuroimage* 14:439–453.
- Chao LL, Haxby JV, Martin A. 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2:913–919.
- Chun MM. 2003. Scene perception and memory. In: Irwin DE, Ross B, editors. *Psychology of learning and motivation: advances in research and theory: cognitive vision*. San Diego, CA: Academic Press. p 79–108.
- Chun MM, Jiang Y. 1998. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognit Psychol* 36:28–71.
- Cox D, Meyers E, Sinha P. 2004. Contextually evoked object-specific responses in human visual cortex. *Science* 304:115–117.
- Davenport JL, Potter MC. 2004. Scene consistency in object and background perception. *Psychol Sci* 15:559–564.
- Diamond R, Carey S. 1986. Why faces are and are not special: an effect of expertise. *J Exp Psychol Gen* 115:107–117.
- Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N. 2006. Domain specificity in visual cortex. *Cereb Cortex* 16:1453–61.
- Epstein R, DeYoe EA, Press DZ, Rosen AC, Kanwisher N. 2001. Neuropsychological evidence for a topographical learning mechanism in parahippocampal cortex. *Cogn Neuropsychol* 18:481–508.
- Epstein R, Graham KS, Downing PE. 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.
- Epstein RA. 2005. The cortical basis of visual scene processing. *Vis Cogn* 12:954–978.
- Epstein RA, Higgins JS, Thompson-Schill SL. 2005. Learning places from views: variation in scene processing as a function of experience and navigational ability. *J Cogn Neurosci* 17:73–83.
- Frith U, Frith CD. 2003. Development and neurophysiology of mentalizing. *Philos Trans R Soc Lond B Biol Sci* 358:459–473.
- Gauthier I. 2000. What constrains the organization of the ventral temporal cortex? *Trends Cogn Sci* 4:1–2.
- Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. 1999. Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects. *Nat Neurosci* 2:568–573.
- Grill-Spector K. 2003. The neural basis of object perception. *Curr Opin Neurobiol* 13:159–166.
- Habib M, Sirigu A. 1987. Pure topographical disorientation—a definition and anatomical basis. *Cortex* 23:73–85.
- Hasson U, Harel M, Levy I, Malach R. 2003. Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37:1027–1041.
- Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn Sci* 4:223–233.
- Henderson JM, Hollingworth A. 1999. High-level scene perception. *Annu Rev Psychol* 50:243–271.
- Hollingworth A, Henderson JM. 1998. Does consistent scene context facilitate object perception? *J Exp Psychol Gen* 127:398–415.
- Ino T, Inoue Y, Kage M, Hirose S, Kimura T, Fukuyama H. 2002. Mental navigation in humans is processed in the anterior bank of the parieto-occipital sulcus. *Neurosci Lett* 322:182–186.
- Intraub H. 1997. The representation of visual scenes. *Trends Cogn Sci* 1:217–222.
- Ishai A, Ungerleider LG, Haxby JV. 2000. Distributed neural systems for the generation of visual images. *Neuron* 28:979–990.
- Janzen G, van Turennout M. 2004. Selective neural representation of objects relevant for navigation. *Nat Neurosci* 7:673–677.
- Kable JW, Chatterjee A. The specificity of action representations in lateral occipitotemporal cortex. *J Cogn Neurosci*. Forthcoming.
- Kanwisher N. 2004. The ventral visual object pathway in humans: evidence from fMRI. In: Chalupa LM, Werner JS, editors. *The visual neurosciences*. Cambridge MA: MIT Press. p 1179–1189.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Katayama K, Takahashi N, Ogawara K, Hattori T. 1999. Pure topographical disorientation due to right posterior cingulate lesion. *Cortex* 35:279–282.
- 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.
- Kosslyn SM, Ganis G, Thompson WL. 2001. Neural foundations of imagery. *Nat Rev Neurosci* 2:635–642.
- Kuipers B. 2000. The spatial semantic hierarchy. *Artif Intell* 119: 191–233.
- Kuipers B, Modayil J, Beeson P, MacMahon M, Savelli F. 2004. Local metrical and global topological maps in the hybrid spatial semantic hierarchy. *IEEE International Conference on Robotics and Automation*.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001. Center-periphery organization of human object areas. *Nat Neurosci* 4:533–539.
- Logothetis NK, Sheinberg DL. 1996. Visual object recognition. *Annu Rev Neurosci* 19:577–621.
- Maguire EA. 2001. The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand J Psychol* 42:225–238.
- Maguire EA, Burgess N, Donnett JG, Frackowiak RSJ, Frith CD, O’Keefe J. 1998. Knowing where and getting there: a human navigation network. *Science* 280:921–924.
- Maguire EA, Frackowiak RSJ, Frith CD. 1997. Recalling routes around London: activation of the right hippocampus in taxi drivers. *J Neurosci* 17:7103–7110.
- Maljkovic V, Martini P. 2005. Short-term memory for scenes with affective content. *J Vis* 5:215–229.
- Martin A, Wiggs CL, Ungerleider LG, Haxby JV. 1996. Neural correlates of category-specific knowledge. *Nature* 379:649–652.
- McCarthy G, Puce A, Gore JC, Allison T. 1997. Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci* 9:605–610.
- McNamara TP, Rump B, Werner S. 2003. Egocentric and geocentric frames of reference in memory of large-scale space. *Psychon Bull Rev* 10:589–595.
- Mendez MF, Cherrier MM. 2003. Agnosia for scenes in topographagnosia. *Neuropsychologia* 41:1387–1395.
- Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 15:1–25.
- O’Craven KM, Kanwisher N. 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J Cogn Neurosci* 12:1013–1023.
- Oliva A, Torralba A. 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int J Comp Vis* 42:145–175.
- Peelen MV, Downing PE. 2005. Selectivity for the human body in the fusiform gyrus. *J Neurophysiol* 93:603–608.
- Pelphrey KA, Morris JP, McCarthy G. 2004. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J Cogn Neurosci* 16:1706–1716.
- Potter MC. 1975. Meaning in visual search. *Science* 187:965–966.
- Puce A, Allison T, Asgari M, Gore JC, McCarthy G. 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci* 16:5205–5215.
- Puce A, Allison T, Bentin S, Gore JC, McCarthy G. 1998. Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci* 18:2188–2199.
- Renninger LW, Malik J. 2004. When is scene identification just texture recognition? *Vision Res* 44:2301–2311.
- Rosch E, Mervis CB, Gray WD, Johnson DM, Boyesbraem P. 1976. Basic objects in natural categories. *Cogn Psychol* 8:382–439.
- Rosenbaum RS, Priselac S, Kohler S, Black SE, Gao FQ, Nadel L, Moscovitch M. 2000. Remote spatial memory in an amnesic person with extensive bilateral hippocampal lesions. *Nat Neurosci* 3: 1044–1048.
- Rosenbaum RS, Ziegler M, Winocur G, Grady CL, Moscovitch M. 2004. “I have often walked down this street before”: fMRI studies on the hippocampus and other structures during mental navigation of an old environment. *Hippocampus* 14:826–835.
- Rousselet GA, Joubert OR, Fabre-Thorpe M. 2005. How long to get the “gist” of real-world natural scenes? *Vis Cogn* 12:852–877.
- Saxe R, Xiao D, Kovacs G, Perrett DI, Kanwisher N. 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42:1435–1446.
- Schyns PG, Oliva A. 1994. From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychol Sci* 5:195–200.
- Shelton AL, Gabrieli JD. 2002. Neural correlates of encoding space from route and survey perspectives. *J Neurosci* 22:2711–2717.
- Somers DC, Dale AM, Seiffert AE, Tootell RBH. 1999. Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc Natl Acad Sci USA* 96:1663–1668.
- Steeves JK, Humphrey GK, Culham JC, Menon RS, Milner AD, Goodale MA. 2004. Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *J Cogn Neurosci* 16:955–965.
- Sugiura M, Shah NJ, Zilles K, Fink GR. 2005. Cortical representations of personally familiar objects and places: functional organization of the human posterior cingulate cortex. *J Cogn Neurosci* 17: 183–198.

- Takahashi N, Kawamura M, Shiota J, Kasahata N, Hirayama K. 1997. Pure topographic disorientation due to right retrosplenial lesion. *Neurology* 49:464-469.
- Tanaka K. 1993. Neuronal mechanisms of object recognition. *Science* 262:685-686.
- Teng E, Squire LR. 1999. Memory for places learned long ago is intact after hippocampal damage. *Nature* 400:675-677.
- Torralba A, Oliva A. 2003. Statistics of natural image categories. *Neural Syst* 14:391-412.
- Tversky B, Hemenway K. 1983. Categories of environmental scenes. *Cognit Psychol* 15:121-149.
- Wolbers T, Busch C. 2005. Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *J Neurosci* 25:3333-3340.