

Outside Looking In: Landmark Generalization in the Human Navigational System

Steven A. Marchette,¹ Lindsay K. Vass,¹ Jack Ryan,¹ and Russell A. Epstein¹

Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

The use of landmarks is central to many navigational strategies. Here we use multivoxel pattern analysis of fMRI data to understand how landmarks are coded in the human brain. Subjects were scanned while viewing the interiors and exteriors of campus buildings. Despite their visual dissimilarity, interiors and exteriors corresponding to the same building elicited similar activity patterns in the parahippocampal place area (PPA), retrosplenial complex (RSC), and occipital place area (OPA), three regions known to respond strongly to scenes and buildings. Generalization across stimuli depended on knowing the correspondences among them in the PPA but not in the other two regions, suggesting that the PPA is the key region involved in learning the different perceptual instantiations of a landmark. In contrast, generalization depended on the ability to freely retrieve information from memory in RSC, and it did not depend on familiarity or cognitive task in OPA. Together, these results suggest a tripartite division of labor, whereby PPA codes landmark identity, RSC retrieves spatial or conceptual information associated with landmarks, and OPA processes visual features that are important for landmark recognition.

Key words: functional magnetic resonance imaging; multivoxel pattern analysis; object recognition; parahippocampal place area; retrosplenial complex; spatial memory

Significance Statement

A central element of spatial navigation is the ability to recognize the landmarks that mark different places in the world. However, little is known about how the brain performs this function. Here we show that the parahippocampal place area (PPA), a region in human occipitotemporal cortex, exhibits key features of a landmark recognition mechanism. Specifically, the PPA treats different perceptual instantiations of the same landmark as representationally similar, but only when subjects have enough experience to know the correspondences among the stimuli. We also identify two other brain regions that exhibit landmark generalization, but with less sensitivity to familiarity. These results elucidate the brain networks involved in the learning and recognition of navigational landmarks.

Introduction

Landmarks are entities that have a special status in navigation because they are associated with specific locations or directions in the world (Lynch, 1960; Siegel and White, 1975; Gallistel, 1990). They can come in many different varieties, including buildings, statues, the shape of a room, or the topography of a natural landscape (Epstein and Vass, 2014). Because of their centrality to many navigational strategies, it is reasonable to hypothesize that

the brain might contain a mechanism for learning and recognizing landmarks. However, a neural locus for such a mechanism has not been clearly demonstrated. Here we use multivoxel pattern analysis (MVPA) of fMRI data to resolve this issue.

A key feature of any putative landmark recognition mechanism would be the ability to associate the different perceptual features that indicate a specific place, treating these features as equivalent, even if they are perceptually distinct. That is, a landmark recognition mechanism should discriminate between stimuli shown in different places but generalize across stimuli encountered in the same place (especially if the same-place stimuli are different views of the same underlying landmark object). To test for this pattern of landmark generalization, we scanned subjects while they viewed the interiors and exteriors of buildings at the University of Pennsylvania (Penn) campus (Fig. 1). Although the façade of a building and a room inside are visually dissimilar from each other, they are both views of the same landmark, with similar navigational significance. Thus, we reasoned

Received June 12, 2015; revised Sept. 25, 2015; accepted Sept. 30, 2015.

Author contributions: S.A.M., L.K.V., and R.A.E. designed research; S.A.M. and J.R. performed research; S.A.M., L.K.V., and R.A.E. analyzed data; S.A.M. and R.A.E. wrote the paper.

This work was supported by National Institutes of Health Grant R01-EY022350 (R.A.E.) and National Science Foundation Grant SBE-0541957. We thank Anthony Stigliani and Nicole Paul for assistance with data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to Steven A. Marchette, Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104. E-mail: stmar@sas.upenn.edu.

DOI:10.1523/JNEUROSCI.2270-15.2015

Copyright © 2015 the authors 0270-6474/15/3514896-13\$15.00/0

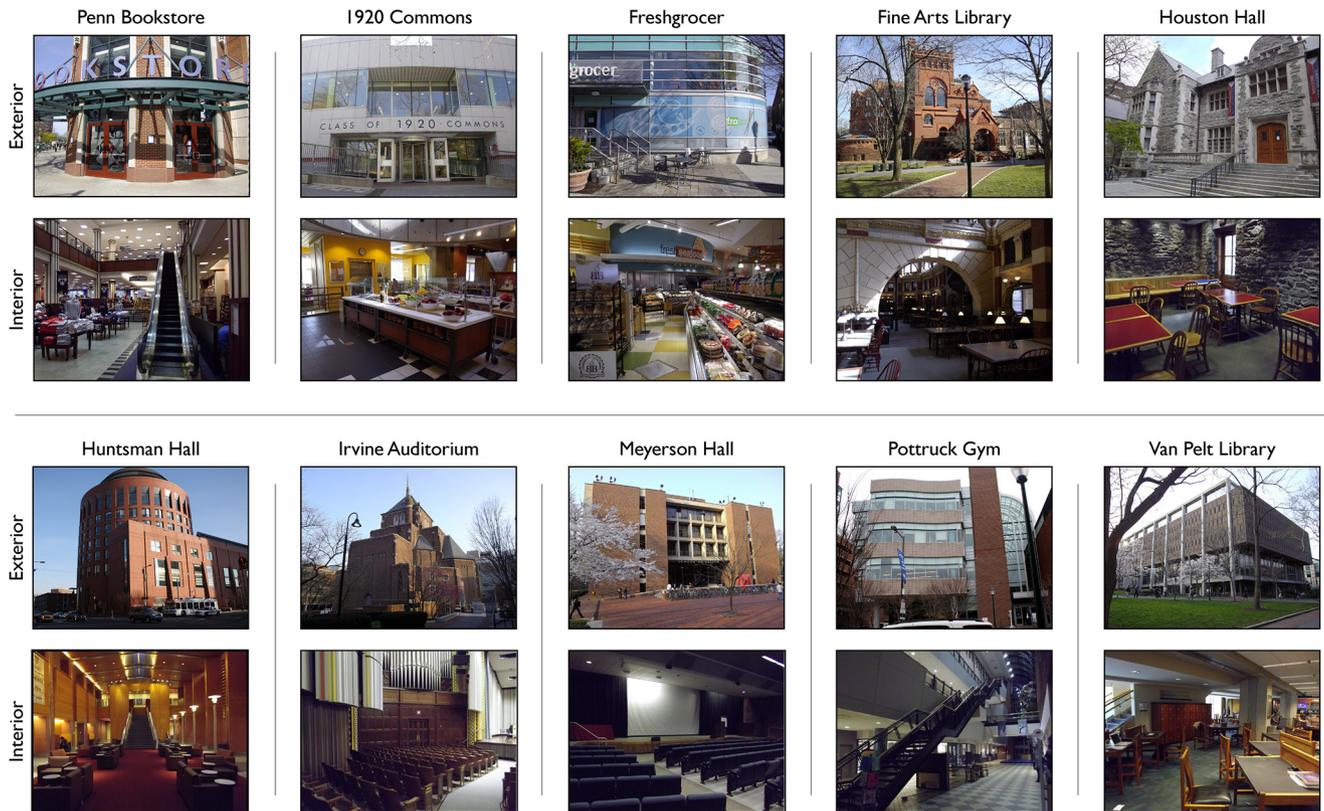


Figure 1. Examples of stimuli. Subjects viewed photographs of the exteriors and interiors of 10 landmark buildings from the University of Pennsylvania campus. One example photograph for each interior and each exterior is shown.

that brain regions involved in landmark recognition should exhibit multivoxel activation patterns that are similar for the interior and exterior of the same building, but dissimilar for the interior of one building and the exterior of another. We further predicted that this generalization across interiors and exteriors should only occur when subjects knew the correspondences among these stimuli, and it should proceed automatically even in the absence of an explicit act of memory retrieval. The combination of these effects would indicate the existence of an abstract representation of landmark identity that would be essential for solving many navigational problems (e.g., figuring out how to use one’s cognitive map of campus to get from an interior space in one building to an interior space in another).

We hypothesized that the parahippocampal place area (PPA), a region at the boundaries of posterior parahippocampal, lingual, and fusiform gyri, would be the brain region that exhibits these predicted patterns. This hypothesis was based on two lines of previous research. First, neuropsychological work indicates that damage to the parahippocampal/lingual region leads to a deficit in landmark recognition (Aguirre and D’Esposito, 1999). Second, neuroimaging work indicates the PPA responds strongly to objects that are suitable as landmarks (Troiani et al., 2012) because they are large in size (Cate et al., 2011; Konkle and Oliva, 2012), distant from the viewer (Amit et al., 2012), located at a navigationally relevant location (Janzen and van Turennout, 2004; Schinazi and Epstein, 2010), associated with a context (Bar and Aminoff, 2003), or definitional of the space around them (Mullally and Maguire, 2011).

We tested the role of the PPA in landmark coding in three experiments by investigating whether PPA exhibits landmark coding that generalizes across interior and exterior views (Exper-

iment 1), whether this generalization requires knowledge of the landmarks’ identity and the correspondences among images (Experiment 2), and whether this generalization is affected by varying the memory retrieval demands placed on the subject (Experiment 3). To anticipate, our results indicate that PPA represents landmark identity in an abstract manner that involves generalization across different stimuli and also displays the other characteristics expected of a landmark recognition mechanism. In addition, two other regions implicated in scene perception and navigation, the retrosplenial complex (RSC) and occipital place area (OPA), exhibited some but not all of these properties, indicating that these regions are also involved in landmark processing but with functional roles that are distinct from the PPA.

Materials and Methods

Participants

Sixteen subjects (8 female; mean age, 20.5 ± 0.8 years) were recruited from the Penn community to participate in Experiment 1, 16 subjects (8 female; mean age, 21.5 ± 1.4 years) were recruited from the Temple University community to participate in Experiment 2, and 24 subjects (12 female; mean age, 21.1 ± 1.2 years) were recruited from the Penn community to participate in Experiment 3. Subjects in Experiments 1 and 3 had at least 2 years of experience with the Penn campus; subjects in Experiment 2 had no or minimal experience with this environment but were matched on years at college. All 56 subjects were healthy, were right-handed, had normal or corrected-to-normal vision, and provided written informed consent in compliance with procedures approved by the University of Pennsylvania Institutional Review Board. Data from five additional subjects were collected but discarded before analysis: three in Experiment 1 (one for neurological abnormality, one for scanner artifact, and one who reported not paying attention to the images during the experiment) and two in Experiment 3 (for excessive head motion).

To ensure that Penn subjects in Experiments 1 and 3 were familiar with the Penn buildings that would be viewed in the scanner, prospective participants were brought in for a prescreening session in which they viewed exterior and interior images of the 10 buildings used in the experiment and seven filler items. For each building, they were asked to select the appropriate name from a list and rate (on a 1–5 scale) their confidence in their answer, their familiarity with the building, and their knowledge of its location. Subjects were only asked to participate in the experiment if they accurately named the interiors and exteriors of all 10 buildings used in the imaging experiment and rated their confidence, familiarity, and location knowledge for each as 3, 4, or 5, with no more than one item rated as a 3. No feedback was given during the prescreening, and images used in the prescreening were not reused in the subsequent experiment. To ensure that Temple students (Experiment 2) were not familiar with the Penn buildings, prospective subjects filled out a web form to determine eligibility. Subjects were asked to judge how many times a week they visited the University of Pennsylvania campus on a scale from 0 (never) to 7 (everyday) and their familiarity with the campus on a 1 (not at all) to 5 (very) scale. Subjects were only asked to participate in the experiment if they visited the Penn campus zero times a week and also rated their overall familiarity as 1 or 2.

In addition to the imaging experiments, we ran two behavioral experiments on Amazon's Mechanical Turk (MTurk). One hundred thirty-seven subjects participated in the first MTurk experiment, and 358 subjects participated in the second experiment. An additional 188 MTurk subjects provided behavioral ratings that contributed to stimulus creation for the second behavioral experiment. All subjects were required to have the Master Worker qualification.

MRI acquisition

Scanning was performed at the Hospital of the University of Pennsylvania using a 3T Siemens Trio scanner equipped with a 32-channel head coil. High-resolution T1-weighted images for anatomical localization were acquired using a three-dimensional magnetization-prepared rapid acquisition gradient echo pulse sequence [repetition time (TR), 1620 ms; echo time (TE), 3.09 ms; inversion time, 950 ms; voxel size, $1 \times 1 \times 1$ mm; matrix size, $192 \times 256 \times 160$]. T2*-weighted images sensitive to blood oxygenation level-dependent contrasts were acquired using a gradient echo echoplanar pulse sequence (TR, 3000 ms; TE, 30 ms; flip angle, 90° ; voxel size, $3 \times 3 \times 3$ mm; field of view, 192; matrix size, $64 \times 64 \times 44$). Visual stimuli were displayed by rear-projecting them onto a Mylar screen at 1024×768 pixel resolution with an Epson 8100 3-LCD projector equipped with a Buhl long-throw lens. Subjects viewed the images through a mirror attached to the head coil. Images subtended a visual angle of $\sim 22.9 \times 17.4^\circ$.

Design and task: fMRI experiments

Experiment 1. To determine fMRI response to different perceptual instantiations of familiar landmarks, Penn subjects were scanned while viewing 440 digital color photographs of interiors and exteriors of Penn campus buildings, shown one at a time. Specifically, for each of 10 prominent campus buildings, subjects viewed 22 images of the exterior facade and 22 images taken within one interior room. Images were presented for 1000 ms each, followed by a 2000 ms gap before the presentation of the next stimulus. To ensure attention to the stimuli, subjects were instructed to press a button as quickly as possible once they recognized the building depicted in each photograph. This task queried subjects' familiarity with each building; they were not asked to retrieve locations or names.

Testing sessions were divided into four scan runs, each of which consisted of 110 stimulus trials and 11 null trials during which the subject viewed a blank screen for 6 s and made no response (total length: 7 min, 18 s per scan run). Subjects viewed interior images of all 10 buildings in two scan runs and exterior images of all 10 buildings in two scan runs. Exterior (E) and interior (I) runs alternated (e.g., E, I, E, I) with the order counterbalanced across subjects. Trials within each scan run were ordered according to a continuous carryover sequence (Aguirre, 2007) so that each building preceded and followed every other building, including itself, exactly once. The specific images used on each trial within a run

were drawn at random from the larger set with the constraint that images did not repeat within a run. A unique carryover sequence was used for each scan run in the experiment.

Experiment 2. In the second experiment, we examined the effect of familiarity on landmark coding, by showing the same Penn landmarks used in Experiment 1 to Temple University students who were unfamiliar with the landmarks. The procedure was mostly identical to Experiment 1, with the following exceptions. Subjects performed the same familiarity judgment task as the subjects in Experiment 1, but in this case, they were instructed to press one button if they recognized the landmark shown on each trial and another button if they did not recognize it. To ensure that these subjects did not become frustrated while attempting to recognize unfamiliar stimuli, we inserted catch trials in which buildings from the Temple campus were shown (interior or exterior views, depending on the format of the Penn buildings shown within the same run). The addition of this catch condition to the continuous-carryover sequence required lengthening each scan run to 144 stimulus trials (12 for each Penn building, plus 12 images of buildings from the Temple campus) and 12 null trials, for a total length of 8 min and 24 s. The specific images used on each trial were drawn at random from the larger set of 480 Penn and 48 Temple photographs with the constraint that images did not repeat over the course of the experiment.

To determine the extent to which Temple students were able to learn the correspondences between the interiors and exteriors of Penn buildings over the course of the experiment, we performed a postscan test. On each trial, subjects were presented with one image of a Penn interior or exterior and asked to pick the image corresponding to the same building from 10 possible choices shown as images in the opposite format. All images were randomly selected from the stimulus set used in the imaging experiment. We then compared their performance with that of naive subjects on Mechanical Turk who did not participate in the fMRI experiment (see below, Design and task: MTurk experiments). We reasoned that if the Temple students outperformed the Mechanical Turk subjects, this would indicate that they had learned some of the correspondences between the interiors and exteriors based on experiencing many different images of them over the course of the experiment.

Experiment 3. In the third experiment, we tested the susceptibility of landmark codes to cognitive interruption, by showing the same Penn landmarks used in the previous experiments to Penn subjects while they performed a concurrent memory task that interfered with memory recall and mental imagery. The design was mostly identical to that of Experiment 1, with one major exception: the subjects in this case learned associations between the Penn landmarks and faces in a prescan training session, and during the scan session, they performed a memory retrieval task on these associations.

To learn these face–place associations, subjects were brought into the laboratory 1 day before scanning and performed 12 alternating phases of study and test. Each of the 10 interiors and 10 exteriors was associated with the face of an unfamiliar person, with the interior and exterior of each building always associated with faces of the opposite gender. In study phases, subjects viewed these 20 scene–face image pairs presented on the screen for 5 s each in a random order and were asked to remember the association between members of the pair. In test phases, a scene and a face were presented on the screen, and subjects had to respond whether these were associated with each other or not; after every trial, the screen flashed red or green to provide feedback on whether the association was accurately remembered. In each test phase, every scene was shown twice: once paired with its correct associate and once paired with an incorrect associate of the same gender. After 12 study–test iterations, subjects were presented with the images of the 10 interiors and 10 exteriors used in training for 1 s each with a 2 s interstimulus interval and asked to mentally imagine the face that was associated with each. Finally, to ensure the associations were well learned, subjects were given a refresher session consisting of six study–test iterations, including feedback, immediately before scanning. Photographs of the Penn buildings used in the study session were different from the photographs shown during imaging.

During the scan session, images of Penn landmarks were shown using the same timing and sequencing parameters as in Experiment 1, with the additional constraint that individual images did not repeat over the

course of the experiment. However, in this case, subjects were instructed to recall the face associated with each interior or exterior and indicate whether the face was male or female by pressing one of two buttons. Thus, this task required landmark recognition insofar as each item had to be identified (e.g., the gym's interior); however, recall or mental imagery of the complementary version of that item (e.g., the gym's exterior) was explicitly discouraged by the fact that the other face associated with the building would imply the opposite response.

Functional localizer. All subjects completed two functional localizer scans at the end of each scan session. These scans were 5 min, 32 s in length, during which subjects performed a one-back repetition detection task on scenes, objects, and scrambled objects presented in 16 s blocks with each stimulus shown for 600 ms each with a 400 ms interstimulus interval.

Design and task: MTurk experiments

In addition to the three scanning experiments, we ran two additional behavioral experiments on MTurk subjects. These experiments tested whether our stimuli contained cues that could support landmark generalization in naive subjects who had no experience with the Penn buildings.

The first MTurk experiment assessed the ability of naive subjects to assess the correspondences between the interiors and exteriors of the Penn buildings using the same task that Temple subjects performed during debriefing. On each trial, subjects were presented with one image of a Penn interior or exterior and asked to pick the image corresponding to the same building from 10 possible choices shown as images in the opposite format. All images were drawn from the stimulus set used in the imaging experiment. Performance was measured by the mean number of correspondences correctly guessed overall.

The second MTurk experiment tested whether interiors and exteriors corresponding to the same Penn building elicited similar conceptual information about the category of the depicted place. To determine an appropriate set of place-category labels for our stimuli, we had 188 MTurk subjects view images of the building interiors and type the name they would use to describe the place. From these responses, we created a list of 32 place categories by taking the five most frequent names given to each interior and removing close synonyms or nonspecific building attributes (e.g., hallway). We then had a different group of 358 MTurk subjects apply these place-category labels to images of the landmarks. On each of 32 trials, subjects read the name of a place category and selected the best exemplar of that category from among images of the 10 landmarks. Approximately half (180) of the subjects viewed only the exteriors; the others (178) viewed only the interiors. We then represented each interior and exterior as a vector that indicated the frequency with which that scene was rated as the best example of each category and measured the conceptual similarities among the interior and exterior scenes by calculating the correlations among their respective place-category vectors. These correlations were then used to test whether interiors and exteriors corresponding to the same landmark received more similar place-category judgments than images corresponding to different landmarks.

fMRI data analysis

Data preprocessing. Functional images were corrected for differences in slice timing by resampling slices in time to match the first slice of each volume. Images were then realigned to the first volume of the scan run, and subsequent analyses were performed within the subjects' own space. Motion correction was performed using MCFLIRT (Jenkinson et al., 2002). Data from the functional localizer scan were smoothed with a 6 mm full-width at half-maximum Gaussian filter; data from the main experiment were not smoothed.

Regions of interest. We identified three scene-selective regions of interest (ROIs) using data from the functional localizer scans: the PPA, RSC, and OPA. These ROIs were defined for each subject individually using a contrast of scenes > objects and a group-based anatomical constraint of scene-selective activation derived from a large number (42) of localizer subjects in our laboratory (Julian et al., 2012). Specifically, each ROI was defined as the top 100 voxels in each hemisphere that responded more to

scenes than to objects and fell within the group-parcel mask for the ROI. This method ensures that all three scene-selective ROIs could be defined in both hemispheres in every subject and that all ROIs contain the same number of voxels, thus facilitating comparisons between regions. We observed similar results when ROIs were defined as all voxels with a localizer contrast significant at $p < 0.001$ uncorrected.

In addition to scene-selective regions, early visual cortex (EVC) was defined based on a contrast of scrambled objects > objects in the functional localizer data. Anatomical ROIs were defined in the hippocampus and presubiculum using the automatic segmentation protocol in FreeSurfer 5.1 (Van Leemput et al., 2009) and in parahippocampal cortex (PHC) based on manual parcellation of the T1-weighted image according to established protocols (Insausti et al., 1998; Pruessner et al., 2002).

Multivoxel pattern analysis. To test the information about landmark identity within each ROI in each subject, we calculated the similarities across scan runs between the multivoxel activity patterns elicited by the 10 interiors and 10 exteriors. If a region contains information about building identity, then patterns corresponding to the same building in different scan runs should be more similar than patterns corresponding to different buildings (Haxby et al., 2001). Moreover, if this effect is observed for patterns elicited by images of different formats (i.e., interior–exterior), then this implies that the landmark identity code generalizes across formats.

To define activity patterns, we used general linear models (GLMs), implemented in FSL (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>), to estimate the response of each voxel to each stimulus condition in each scan run. Each runwise GLM included one regressor for each building (10 total), regressors for motion parameters, and nuisance regressors to exclude outlier volumes discovered using the Artifact Detection Toolbox (http://www.nitrc.org/projects/artifact_detect/). An additional nuisance regressor was included in Experiment 2 to model response to the Temple buildings. High-pass filters were used to remove low temporal frequencies before fitting the GLM, and the first three volumes of each run were discarded to ensure data quality. Multivoxel patterns for each ROI were then created by concatenating the estimated responses across all voxels in both hemispheres.

To determine similarities between activity patterns, we calculated Pearson correlations between patterns in different scan runs. Individual patterns were normalized before this computation by subtracting the grand mean pattern (i.e., the cocktail mean) for each run (Vass and Epstein, 2013). We then computed three discrimination scores based on these correlation values, each of which involved comparing the mean correlation across scan runs for patterns corresponding to the same landmark with the mean correlation across scan runs for patterns corresponding to different landmarks. First, to test for coding of information about building exteriors, we performed this calculation for patterns elicited by exteriors (“exterior decoding”). Second, to test for coding of information about building interiors, we performed this calculation for patterns elicited by interiors (“interior decoding”). Finally, to test for coding of landmark identity that generalizes across format, we compared the average correlation between exterior and interior patterns corresponding to different buildings with the average correlation between exterior and interior patterns corresponding to the same building (“cross-decoding”). The exterior and interior discriminations score were each based on comparisons between one pair of scan runs (e.g., runs 1–3 exterior, 2–4 interior), whereas the cross-decoding discrimination score was based on comparisons between all four pairings of different-format scan runs (i.e., runs 1–2, 2–3, 3–4, 1–4).

Permutation tests were used to determine chance-level performance for each type of discrimination score (exterior, interior, and cross). For each type of discrimination, we independently shuffled the condition labels in the runs being compared and recalculated the mean discrimination score observed across participants for that permutation. We performed this procedure 10,000 times per experiment for each of the functional ROIs (PPA, RSC, OPA, and EVC). In all cases, the mean chance decoding was 0.

Changes over the course of the experiment. To test whether the ability to cross-decode might change over the course of the experiment (e.g., because of landmark learning in naive subjects in Experiment 2), we per-

formed an additional analysis in which we calculated landmark decoding in the first half (runs 1–2) and second half (3–4) separately. We then tested whether cross-decoding was significant in each half and whether the cross-decoding index changed from the first to the second half of the experiment.

Searchlight analysis. To test for cross-decoding of landmark identity outside of our predefined ROIs, we implemented a whole-brain searchlight analysis (Kriegeskorte et al., 2006) in which we centered a small spherical ROI (radius, 5 mm) around every voxel of the brain, calculated the landmark discrimination within this spherical neighborhood using the method described above, and assigned the resulting value to the central voxel. Searchlight maps from individual subjects were then aligned to the Montreal Neurological Institute (MNI) template with a linear transformation and submitted to a second-level random-effects analysis to test the reliability of discrimination across subjects. To find the true type I error rate, we performed Monte Carlo simulations that permuted the sign of the whole-brain maps from individual subjects (Nichols and Holmes, 2002). Voxels were considered significant if they survived correction for multiple comparisons across the entire brain.

Comparison of cross-decoding in anterior and posterior PPA. We further explored the distribution of cross-decoding in the PPA based on previous reports of a functional division between anterior and posterior PPA (Baldassano et al., 2013). Each subject's scene-selective PPA (defined, in this case, as voxels exhibiting greater response to scenes than to objects at $p < 0.001$ uncorrected) was divided into an anterior section contained within PHC and a posterior section outside of PHC. Cross-decoding performance was then calculated separately for each section. In addition, within PHC we tested whether cross-decoding was specific to the scene-selective portion (i.e., the anterior PPA) by calculating the correlation between cross-decoding performance for the searchlight surrounding every voxel in PHC against the scene selectivity for that voxel (as defined by the contrast scenes greater than objects in the localizer runs).

Analysis of visual similarities

To determine whether there were commonalities of low-level visual features between interiors and exteriors that might be sufficient to drive cross-decoding, we ran three visual models on the stimuli used in the imaging experiment: pixelwise intensity, the GIST model (Oliva and Torralba, 2001), and HMAX [Riesenhuber and Poggio, 1999; using the implementation by Theriault et al. (2011)]. We then used the output of these models to quantify the physical similarity among the images. For pixelwise intensity, the similarity between images was measured by their pixelwise correlation; for the GIST model, we measured the distance between the GIST descriptors for the images; and for HMAX, we calculated the correlation between the image signatures from the C2 (complex composite, i.e., view-invariant) layer.

We then tested whether any of these visual similarity metrics could discriminate between Penn buildings. To make these comparisons analogous to the MVPA analyses, we computed model similarity between every pair of buildings by calculating the average pairwise similarity between images of those buildings, excluding any comparison of an image to itself. We then used a t test to determine whether model similarity was higher for comparisons corresponding to the same building than for comparisons corresponding to different buildings. We used this method to calculate exterior decoding based on model similarity between exterior images, interior decoding based on model similarity between interior images, and cross-decoding based on model similarity between interiors and exteriors. In addition, to facilitate comparison between model performance and the ability of naive subjects to guess the correspondence between interior and exterior images, we also report descriptive statistics for classification accuracy, as measured by the proportion of images for which the most similar image was also of the same landmark.

Results

Landmark discrimination and generalization in the PPA

Our first goal was to establish whether it was possible to use MVPA to cross-decode between interiors and exteriors in the PPA. (Other brain regions, including RSC and OPA, are considered below.) To this end, in Experiment 1 we scanned 16 Univer-

sity of Pennsylvania students while they viewed images of the interiors and exteriors of Penn buildings. Analysis of multivoxel patterns revealed that exteriors could be decoded from exteriors based on activity patterns within the PPA ($t_{(15)} = 5.041$, $p = 0.00001$), consistent with previous results (Morgan et al., 2011; Epstein and Morgan, 2012), and there was a nearly significant trend toward decoding of interiors from interiors ($t_{(15)} = 2.116$, $p = 0.052$). Critically, exteriors and interiors could also be decoded from each other ($t_{(15)} = 3.110$, $p = 0.0072$); that is, the identities of the patterns elicited when viewing the exteriors could be decoded based on patterns elicited when viewing the interiors, and vice versa (Fig. 2). This cross-decoding suggests that, to at least some extent, the PPA considers the exterior and interior of each building to be the same entity.

One interpretation of these findings is that the PPA supports a high-level representation that abstracts across very different stimuli corresponding to the same landmark. However, an alternative possibility is that the cross-decoding reflects visual similarities between the interior and exterior of each building, which might not be salient to the observer but are picked up on by the PPA. To test this possibility, in Experiment 2 we scanned 16 students from Temple University while they viewed the same set of interior and exterior views of Penn buildings. These subjects were unfamiliar with the Penn campus, a fact that we verified through prescan screening. We reasoned that if the cross-decoding between interiors and exteriors observed in Experiment 1 reflects visual similarities between the images, then it should be observed in the Temple students. In contrast, if cross-decoding reflects an understanding of which interior corresponds to which exterior, then it should not be found in the Temple students, who did not have this knowledge.

Decoding of exteriors from exteriors based on PPA activity patterns was well above chance in these subjects ($t_{(15)} = 4.689$, $p = 0.003$), as was decoding of interiors from interiors ($t_{(15)} = 3.995$, $p = 0.0012$); moreover, comparison across experiments showed that exterior-from-exterior and interior-from-interior decoding was just as strong in Temple students as it was in Penn students (exteriors: $t_{(30)} = -1.153$, $p = 0.258$; interiors: $t_{(30)} = 0.512$, $p = 0.613$). This was expected, given that the exteriors were all visually distinct from each other, as were the interiors. Critically, cross-decoding between exteriors and interiors was at chance in the Temple students ($t_{(15)} = 0.225$, $p = 0.825$), suggesting that this cross-decoding relies on an understanding of which building is which. Direct comparison between the two experiments verified that cross-decoding was significantly reduced in the Temple students compared with the Penn students ($t_{(30)} = -2.310$, $p = 0.028$).

These findings suggest two possibilities. First, the PPA might form a single identity code for each familiar landmark, which can be elicited by either interior or exterior images. Second, the PPA might form separate representations of the exterior view and interior view, but these representations might be linked together, so that viewing an exterior leads to activation of the exterior representation and then subsequent activation of the interior representation (with the opposite causality when viewing an interior view). In essence, the second account attributes cross-decoding to the elicitation of the unseen-format view through mental imagery or memory retrieval of additional information that is associated with both views. These two accounts make different predictions about the relative susceptibility of within-format and across-format decoding to cognitive interruption. Under the first account, there is a single representation elicited by exterior and interior images of each landmark, so any cognitive manipulation

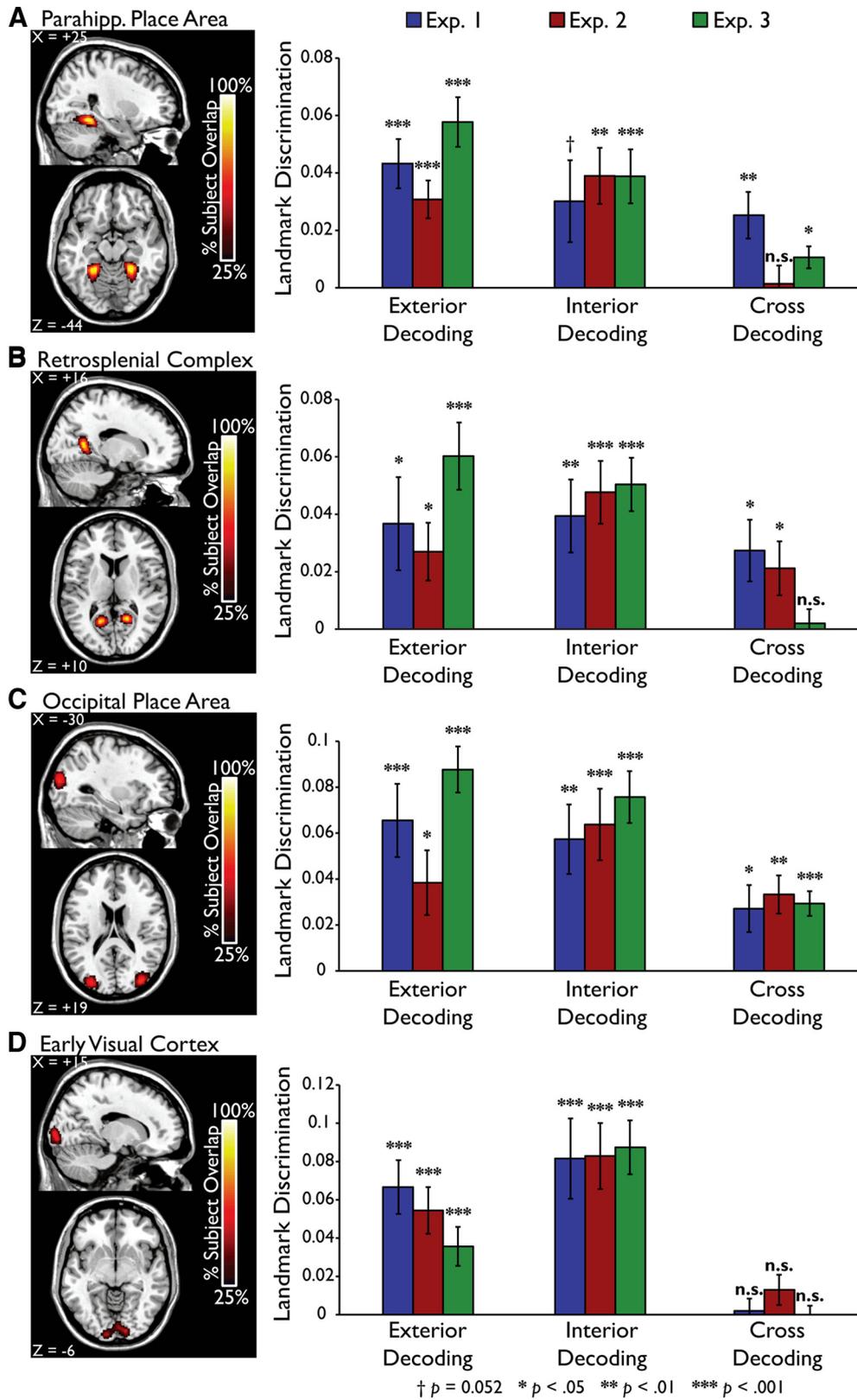


Figure 2. Landmark discrimination in scene-selective regions and EVC. Left, ROI used in the MVPA analysis (A–D). Colors indicate the number of subjects for which each voxel is included in the ROI. Right, Landmark discrimination, defined as greater similarity for fMRI activation patterns corresponding to the same landmark than for fMRI activation patterns corresponding to different landmarks. Every region could discriminate landmark exteriors from other exteriors, and interiors from other interiors, in all three experiments. In PPA, cross-decoding between interiors and exteriors was significant in subjects familiar with the campus (Experiments 1 and 3), but not in subjects unfamiliar with the campus (Experiment 2). Cross-decoding in RSC was significant in subjects familiar with the campus (Experiment 1) and also subjects unfamiliar with the campus (Experiment 2) but was abolished by an interfering memory task (Experiment 3). OPA could cross-decode in all three experiments regardless of familiarity or mnemonic demands, whereas EVC could never reliably cross-decode.

that affects across-format decoding should affect within-format decoding just as strongly. Under the second account, however, it should be possible to selectively reduce across-format decoding by giving subjects a task that interrupts the hypothesized memory retrieval stage.

This logic provided the motivation for Experiment 3. Twenty-four Penn students were scanned using a design that was identical to that of Experiment 1. However, in this case, subjects were first trained to associate a unique human face with each of the interior and exterior views. For each building, the exterior view was paired with the face of one gender, while the interior cue was paired with the face of the opposite gender. During the scan session, rather than simply reporting familiarity with the building, as subjects had done in the previous experiments, participants were asked to imagine the face paired with the view and report its gender. We chose faces because previous work indicates that the PPA does not respond strongly to faces and, to our knowledge, there is no evidence that face identity can be read out from PPA activity patterns. Thus, this task served to interrupt any postrecognition processing in the PPA related to the recall or imagery of associated buildings, without eliciting any competing representations.

Within-format decoding of exteriors from exteriors remained significant in the PPA ($t_{(23)} = 6.674, p = 0.0000008$), as did within-format decoding of interiors from interiors ($t_{(23)} = 4.132, p = 0.0004$); moreover, within-format decoding performance did not differ from Experiment 1 (exteriors: $t_{(38)} = 1.139, p = 0.262$; interiors: $t_{(38)} = 0.530, p = 0.599$). Critically, despite the interfering memory task, interiors and exteriors could be successfully cross-decoded ($t_{(23)} = 2.785, p = 0.011$). Direct comparison between Experiments 1 and 3 found that the cross-decoding effect was marginally reduced in the current experiment compared with the former ($t_{(38)} = -1.812, p = 0.078$). These results are consistent with a scenario under which the exterior and interior of a building elicited a common code in the PPA, whose retrieval was not interrupted by the concurrent task. However, the existence of representational overlap within PPA does not preclude the possibility that separate representations of interiors and exteriors might also exist, and the marginal reduction in decoding from Experiment 1 to Experiment 3 could reflect the elimination of the contribution of these separate yet coactivated representations.

In summary, these results show that the PPA exhibits key characteristics of a landmark recognition mechanism. Multivoxel patterns in the PPA discriminated between landmarks and generalized across different visual instantiations of the same landmark, as evidenced by significant cross-decoding between interiors and exteriors in Experiment 1. Cross-decoding was not significant in Temple subjects who were unfamiliar with the landmarks in Experiment 2 (although see below, Changes across the course of the experiment). In contrast, cross-decoding remained significant in Penn subjects in the presence of a concurrent memory retrieval task in Experiment 3.

Landmark discrimination and generalization in other brain regions

Although our main focus was the PPA, we also examined responses in the two other scene-selective regions, the RSC and OPA. Previous work has indicated that scenes can be classified based on multivoxel patterns in these regions (Walther et al., 2009; Kravitz et al., 2011; Morgan et al., 2011; Epstein and Morgan, 2012), and RSC has been specifically implicated in landmark coding (Auger et al., 2012; Auger and Maguire, 2013). Consistent

with these results, we found that activity patterns in both RSC and OPA allowed classification of exteriors from exteriors, and interiors from interiors, in all three experiments (RSC exterior decoding: Experiment 1, $t_{(15)} = 2.266, p = 0.039$; Experiment 2, $t_{(15)} = 2.685, p = 0.017$; Experiment 3, $t_{(23)} = 5.149, p = 0.00003$; RSC interior decoding: Experiment 1, $t_{(15)} = 3.104, p = 0.0073$; Experiment 2, $t_{(15)} = 4.369, p = 0.0006$; Experiment 3, $t_{(23)} = 5.412, p = 0.00002$; OPA exterior decoding: Experiment 1, $t_{(15)} = 4.112, p = 0.0009$; Experiment 2, $t_{(15)} = 2.719, p = 0.0158$; Experiment 3, $t_{(23)} = 8.710, p = 0.0000001$; OPA interior decoding: Experiment 1, $t_{(15)} = 3.785, p = 0.0018$; Experiment 2, $t_{(15)} = 4.086, p = 0.00097$; Experiment 3, $t_{(23)} = 6.697, p = 0.0000008$).

We also observed significant cross-classification of exteriors from interiors, and interiors from exteriors, in the RSC and OPA (Fig. 2B,C). Notably, the pattern of results across experiments differed from the pattern exhibited by the PPA. In RSC, significant cross-classification was observed in Experiment 1 ($t_{(15)} = 2.547, p = 0.022$) and Experiment 2 ($t_{(15)} = 2.259, p = 0.039$) but not in Experiment 3 ($t_{(23)} = 0.400, p = 0.692$). Thus, cross-decoding in RSC was not significantly reduced in Temple students compared with Penn students (Experiment 1 vs Experiment 2: $t_{(30)} = -0.435, p = 0.667$) but was significantly reduced by the addition of a concurrent memory retrieval task (Experiment 1 vs Exp 3: $t_{(38)} = -2.144, p = 0.044$). In OPA, on the other hand, significant cross-classification was observed in all three experiments (Experiment 1: $t_{(15)} = 2.656, p = 0.018$; Experiment 2: $t_{(15)} = 4.021, p = 0.0011$; Experiment 3: $t_{(23)} = 5.443, p = 0.00002$).

These results suggest that cross-decoding of landmark identity exhibited different profiles of sensitivity to landmark familiarity and cognitive interruption in the three scene-selective regions. In PPA, decoding was possible when the landmarks were familiar (Experiments 1 and 3), but it was reduced when the landmarks were unfamiliar (Experiment 2); for RSC, cross-decoding was possible for both familiar and unfamiliar landmarks (Experiments 1 and 2) but abolished by an interfering memory retrieval task (Experiment 3); and for OPA, cross-decoding was not affected by landmark familiarity or the interfering memory task. To confirm that these patterns of sensitivity represented true differences across ROIs, we submitted cross-decoding performance to a 3×3 mixed-model ANOVA with a within-subjects factor for ROI and a between-subjects factor for experiment. We observed a main effect of ROI ($F_{(2,106)} = 12.132, p = 0.00002$), with greatest cross-decoding in OPA, likely because of its consistency across all three experiments. Critically, we also observed a significant interaction of ROI and experiment ($F_{(4,106)} = 5.098, p = 0.0009$), suggesting a triple dissociation in the contribution of the PPA, RSC, and OPA to landmark identification.

Beyond scene-selective regions, we investigated coding within EVC, defined by a contrast of scrambled objects greater than intact objects, as well as anatomically defined structures within the medial temporal lobe, including presubiculum and the hippocampus. For EVC, we anticipated that activity patterns would be able to decode exteriors from exteriors and interiors from interiors on the basis of the visual similarities among images depicting the same place, and indeed this is what we observed in all three experiments (exteriors: Experiment 1, $t_{(15)} = 5.426, p = 0.00007$; Experiment 2, $t_{(15)} = 4.517, p = 0.0004$; Experiment 3 $t_{(23)} = 3.896, p = 0.0007$; interiors: Experiment 1, $t_{(15)} = 3.891, p = 0.001$; Experiment 2, $t_{(15)} = 4.809, p = 0.0002$; Experiment 3, $t_{(23)} = 6.206, p = 0.000002$). However, it was not possible to use activity patterns in EVC to cross-decode in any of the three experiments (Experiment 1, $t_{(15)} = 0.303, p = 0.766$; Experiment 2,

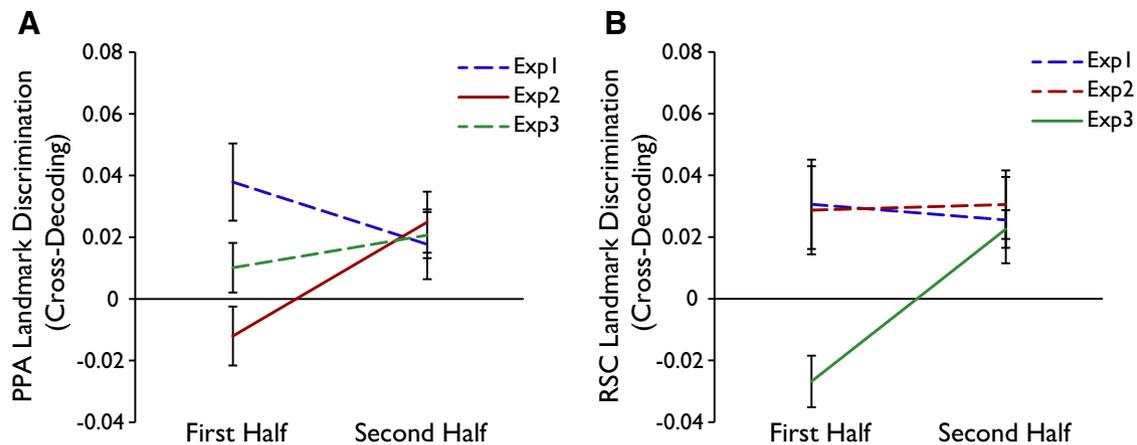


Figure 3. Changes in cross-decoding over the course of the experiment. **A**, Cross-decoding in PPA was stable across the first and second halves of Experiments 1 and 3, in which subjects were Penn students who had real-world experience with the landmarks. In contrast, cross-decoding increased over the course of Experiment 2, in which subjects were Temple students who were initially unfamiliar with the landmarks, suggesting learning of landmarks from visual exposure. **B**, Cross-decoding in RSC was stable across halves in Experiments 1 and 2 but significantly increased over the course of Experiment 3, suggesting that the concurrent memory task became less interfering over time. Solid lines indicate significant changes in cross-decoding; dashed lines are nonsignificant.

$t_{(15)} = 1.639$, $p = 0.112$; Experiment 3, $t_{(23)} = -0.202$, $p = 0.842$). This result is not surprising given the many perceptual dissimilarities between interior and exterior images; indeed, simple visual models were also incapable of cross-decoding (see below, What underlies landmark generalization in naive subjects?). We did not observe decoding of exteriors from exteriors, interiors from interiors, or cross-decoding in the presubiculum (all p values >0.27) or hippocampus (all p values >0.26). In addition, because previous work suggests that retrosplenial cortex proper (BA29/30) might have a role in landmark processing that is distinct from the more posterior portions of RSC located in the parietal-occipital sulcus (Auger et al., 2012), we performed a separate set of analyses on this region. Results for anatomically defined retrosplenial cortex (BA29/30) were identical to results reported above for functionally defined RSC.

Changes over the course of the experiment

The results presented above suggest that cross-decoding in PPA depends on familiarity with the landmarks. But what is the nature of the required familiarity? Does cross-decoding require real-world navigational experience with the buildings, or might landmark representations be built up from visual exposure alone? Although the PPA does not seem to automatically detect the features common to the exterior and interior of a landmark (if it did, then cross-decoding should not be sensitive to familiarity), visual or conceptual commonalities may serve as a basis from which landmark identity could be ascertained from extensive visual exposure. Indeed, behavioral evidence suggested that Temple students learned about the landmarks over the course of the scan session (see below, What underlies landmark generalization in naive subjects?).

To explore whether learning based on visual experience was evident in PPA activity patterns, we calculated cross-classification performance separately for the first half of Experiment 2 (scan runs 1 and 2) and the second half (scan runs 3 and 4; Fig. 3). The key question here is whether cross-decoding increased as the Temple students became familiar with the stimuli. Indeed, we found evidence for landmark learning in the PPA of Temple students. There was a significant increase in cross-decoding performance between the first and second halves of Experiment 2 ($t_{(15)} = 2.526$, $p = 0.023$), and cross-classification

in the second half of the experiment was significant ($t_{(15)} = 2.518$, $p = 0.023$). Thus, it seems that the PPA can build up landmark representations from visual exposure alone, even in subjects who have no real-world experience with the landmarks.

We also performed this analysis for Penn students in Experiments 1 and 3. In this case, we expected cross-decoding to be stable over time, because these subjects came into the experiment with extensive knowledge of the landmarks. As expected, there was no change in the PPA across the first and second halves of the experiment for Penn students in Experiment 1 ($t_{(15)} = 1.149$, $p = 0.269$) or Experiment 3 ($t_{(23)} = -0.878$, $p = 0.389$). To confirm the difference between experiments, we submitted these cross-decoding discrimination scores to a mixed-model ANOVA with a within-subjects factor for experiment half and a between-subjects factor for experiment. This test confirmed a significant interaction of experiment and experiment half ($F_{(2,53)} = 3.377$, $p = 0.042$).

When we performed the same analysis on RSC, we observed a different pattern of results. Here we observed no changes in cross-decoding performance over the course of Experiment 1 ($t_{(15)} = 0.332$, $p = 0.745$) or Experiment 2 ($t_{(15)} = -0.121$, $p = 0.905$) but a quite dramatic increase over the course of Experiment 3 ($t_{(23)} = 4.802$, $p = 0.00008$). The differences between experiments were confirmed by a significant interaction between experiment and experiment half ($F_{(2,53)} = 5.755$, $p = 0.005$). Notably, in the first half of Experiment 3, the multivoxel patterns in RSC elicited by the interior and exterior of each landmark were reliably less similar to each other than the multivoxel patterns elicited by the exterior of one landmark and the interior of another ($t_{(23)} = -3.221$, $p = 0.004$), i.e., the opposite of correct classification. This effect switched signs to become positive, indicating significant cross-classification in the second half of the experiment ($t_{(23)} = 3.690$, $p = 0.001$). Although the reason for this switch is unclear, one possibility is that RSC is heavily involved in the face memory task in the first half of Experiment 3, thus masking the underlying landmark code. Indeed, the representations of interior and exterior may have been driven apart while performing the face task to reduce contamination from the other face associated with the landmark. The task might then be performed in a more automated manner not involving RSC in the second half of the experiment.

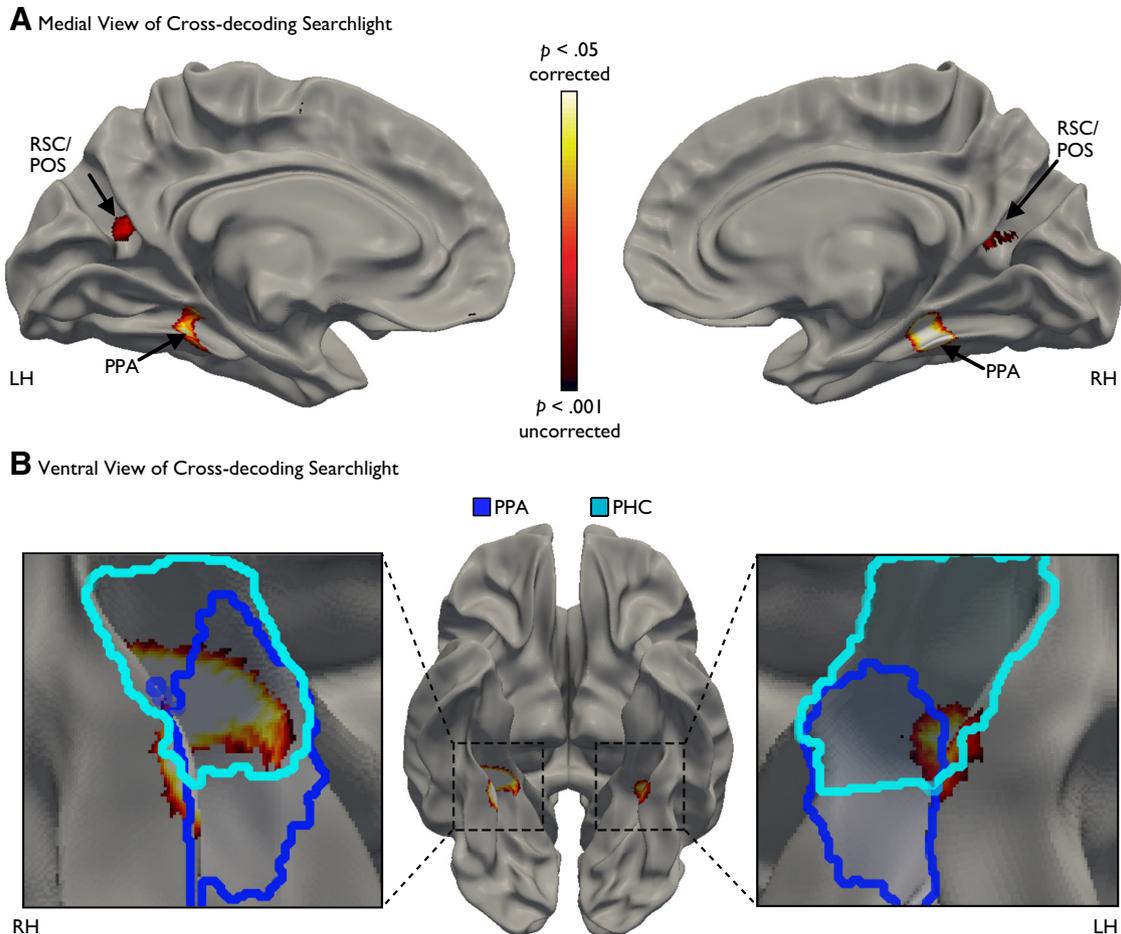


Figure 4. Whole-brain searchlight analysis for cross-decoding in Experiment 1. **A**, Medial view. **B**, Ventral view. Voxels in yellow are significant ($p < 0.05$) after correcting for multiple comparisons across the entire brain; voxels in orange are significant at uncorrected significance levels. Consistent with the results of our ROI analysis, cross-decoding was significant with PPA bilaterally at corrected levels and was significant at a more liberal threshold ($p < 0.001$, uncorrected) within bilateral RSC and right OPA. The outline of PPA and RSC was created by creating a group t statistic in standard space for the contrast scenes greater than objects, thresholded at $p < 0.001$ (corrected). The outline of anatomically defined PHC was manually segmented on the standard space brain.

In contrast to PPA and RSC, we did not observe any changes over the course of Experiments 1, 2, or 3 in OPA (all t values < 1.480 , all p values > 0.16). To confirm that the changes with time differed as a function of region, we submitted the change in cross-decoding from first to second half of each experiment to a 3×3 mixed-model ANOVA with a within-subjects factor for ROI and a between-subjects factor for experiment. We observed no main effect of ROI ($F_{(2,106)} = 0.201$, $p = 0.818$) but a significant interaction between ROI and experiment ($F_{(4,106)} = 5.956$, $p = 0.0002$). This confirms that PPA, RSC, and OPA showed different patterns of change across experiments, with an increase of decoding performance in the PPA in Experiment 2 only, an increase in decoding performance in RSC in Experiment 3 only, and no change in any experiment in OPA.

Whole-brain searchlight analysis

In our final set of fMRI analyses, we used a whole-brain searchlight analysis to identify other regions outside of our ROIs that might be capable of cross-decoding. Results for Experiment 1 are shown in Figure 4. Within the searchlight analysis, significant cross-decoding was limited to bilateral PPA ($p < 0.05$ corrected for multiple comparisons across the entire brain; MNI coordinates: right: 30, -38 , -15 ; left: -30 , -41 , -9). In addition, cross-decoding was observed in bilateral RSC (MNI coordinates:

right: 18, -57 , 18; left: -18 , -63 , 21) and right OPA (MNI coordinates: 40, -74 , 18) at more liberal thresholds ($p < 0.005$ uncorrected). No significant cross-decoding was observed at corrected thresholds in Experiments 2 and 3, although bilateral OPA and right RSC were observed in Experiment 2 at an uncorrected threshold of $p < 0.005$ (MNI coordinates: right OPA: 32, -79 , 30; left OPA: -36 , -90 , 14; right RSC: 12, -56 , 16) and OPA and PPA were observed in Experiment 3 at uncorrected thresholds of $p < 0.005$ and $p < 0.05$, respectively (MNI coordinates: right OPA: 42, -77 , 27; left OPA: -28 , -84 , 24; right PPA: 28, -41 , 18; left PPA: -29 , -44 , -11).

Close inspection of these searchlight results suggested that cross-decoding was found primarily in the anterior portion of the PPA (Fig. 4B). To explore a possible anterior/posterior division, we divided each subject's PPA into an anterior section contained within PHC and a posterior section outside of PHC. We then calculated the cross-decoding ability for each section in the two experiments for which significant PPA cross-decoding was observed (i.e., Experiments 1 and 3; Fig. 5A). A mixed-model ANOVA with anterior section versus posterior section as a within-subject factor and experiment (1 vs 3) as a between-subjects factor revealed a significant main effect of anterior versus posterior ($F_{(1,38)} = 5.386$, $p = 0.026$), with cross-decoding greater in the anterior portion of the PPA within PHC, and a

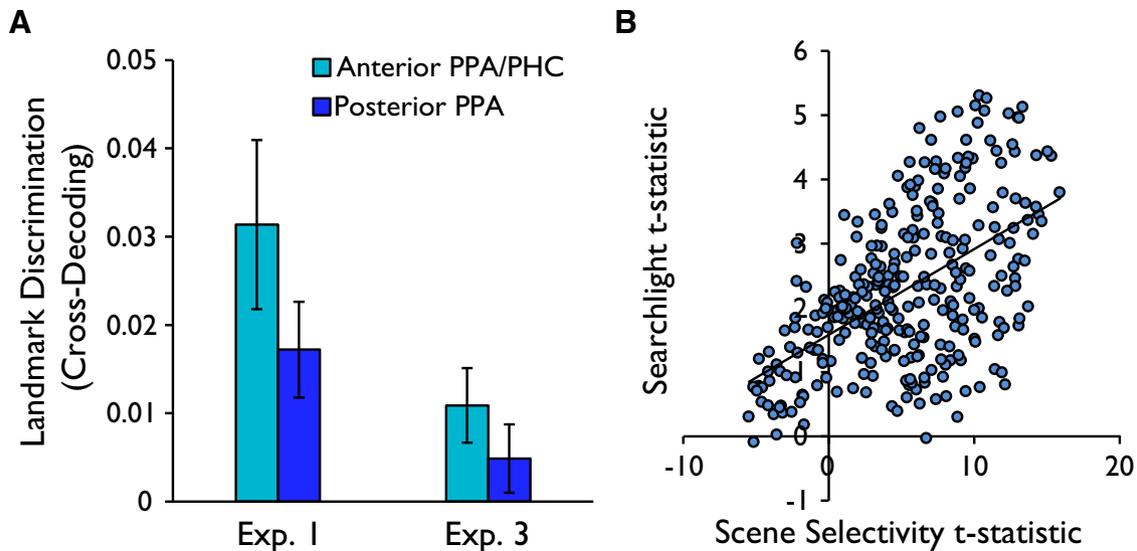


Figure 5. Anterior PPA is the strongest locus of cross-decoding. **A**, Cross-decoding in anterior and posterior PPA of Penn students. Each subject's scene-selective PPA was divided into an anterior section contained within PHC and a posterior section outside of PHC. Cross-decoding performance was greater in the anterior section than in the posterior section, consistent with the result of the searchlight analysis. **B**, Scene selectivity in PHC predicts cross-decoding. The scatterplot depicts the relationship between scene selectivity (x -axis) and cross-decoding (y -axis) for voxels in the PHC. Scene selectivity was measured at each voxel as the group-level t statistic for a contrast of scenes greater than objects, and cross-decoding was measured as the t statistic observed when that voxel served as a searchlight center in the MVPA analysis of cross-decoding. Inspection of the results indicates that cross-decoding is strongest in searchlights surrounding the most scene-selective voxels.

main effect of experiment ($F_{(1,38)} = 5.882$, $p = 0.020$) but no interaction between anterior/posterior and experiment ($F_{(1,38)} = 0.876$, $p = 0.355$). Thus, the strongest cross-decoding was indeed located in the anterior portion of the PPA located within PHC. To determine whether cross-decoding was specific to the scene-selective portion of PHC (i.e., the anterior PPA), for each subject in Experiments 1 and 3 we calculated the correlation between cross-decoding performance for the searchlight surrounding every voxel in PHC against the scene selectivity for that voxel (as defined by the contrast scenes greater than objects in the localizer runs). We observed a correlation that was reliable across subjects ($r = 0.16$, $t_{(39)} = 4.297$, $p = 0.0001$), suggesting that cross-decoding in PHC was primarily found in the scene-selective portion (Fig. 5B).

What underlies landmark generalization in naive subjects?

One notable aspect of the results is that OPA and RSC were capable of cross-decoding in naive subjects in Experiment 2 who did not come into the experiment with knowledge about the correspondence between the interiors and exteriors of the buildings. Moreover, cross-decoding was significant in the PPA in the second half of this experiment. This suggests that there may be visual or conceptual features shared by the interiors and exteriors that allow for cross-decoding, even in the absence of long-term knowledge about which landmark is which.

Indeed, in a postscan test, Temple subjects in Experiment 2 were able to judge some of the correspondences between the interior and exteriors of the Penn buildings (mean correct, 43.4%; chance, 10%; $t_{(15)} = 11.928$, $p = 5.0 \times 10^{-9}$). Moreover, even Mechanical Turk subjects who were viewing the stimuli for the first time were able to determine the correspondences between exteriors and interiors at rates above chance (mean correct, 22.0%; $t_{(136)} = 8.672$, $p = 1.0 \times 10^{-14}$). The difference in performance between the Temple subjects and the Mechanical Turk subjects was significant ($t_{(151)} = -5.140$, $p = 8.0 \times 10^{-7}$), indicating that the within-scan experience

of the Temple subjects led to additional knowledge about interior–exterior correspondences.

What are the features that allow for above-chance cross-decoding in brain regions in naive subjects and above-chance performance on behavioral matching? One possibility is that there are low-level visual features shared by corresponding interiors and exteriors. To test for this possibility, we ran three visual feature models on the images used in the fMRI experiment and tested whether similarity in these models could predict which exteriors and interiors were paired together. These models were pixelwise correlation, the GIST model (Oliva and Torralba, 2001), and the HMAX model (Riesenhuber and Poggio, 1999). All three of the models could accurately classify landmark interiors by comparison with other interiors (pixelwise correlation mean correct, 34.2%; chance, 10%; $t_{(98)} = 3.347$, $p = 0.0012$; GIST mean correct, 65.8%; $t_{(98)} = 3.717$, $p = 0.0003$; HMAX mean correct, 56.7%; $t_{(98)} = 6.747$, $p = 1.0 \times 10^{-9}$), and two of them could accurately classify exteriors based on comparison with other exteriors (GIST mean correct, 68.8%; $t_{(98)} = 4.561$, $p = 0.00001$; HMAX mean correct, 59.2%; $t_{(98)} = 4.450$, $p = 0.00002$) with a marginal trend for the third (pixelwise correlation mean correct, 32.5%; $t_{(98)} = 1.882$, $p = 0.063$). However, none of these models could significantly cross-decode between exteriors and interiors (pixelwise correlation mean correct, 12.7%; $t_{(98)} = 0.5979$, $p = 0.55$; GIST mean correct, 16.3%; $t_{(98)} = 0.5864$, $p = 0.56$; HMAX mean correct, 17.2%; $t_{(98)} = 0.8487$, $p = 0.40$). The failure of these models to cross-decode indicates that low-level visual similarities are insufficient to explain the cross-decoding observed in OPA, PPA, or RSC. Indeed, as one might expect, the results of these models roughly parallel the MVPA results in EVC: successful classification of exteriors from exteriors and interior from interiors, but no cross-decoding.

A second possibility is that naive subjects might be able to guess the correspondence among interiors and exteriors because both evoked the same category of place (e.g., the exterior and

interior of the Penn bookstore both clearly depict a bookstore). We tested the place-category judgments of a group of Amazon MTurk subjects and found that they were significantly more similar for interiors and exteriors corresponding to the same landmark than for those corresponding to different landmarks ($t_{(98)} = 3.150, p = 0.002$). Critically, the subjects who made these judgments only viewed interiors or exteriors but never both, so the similarity in their conceptual judgments could not have been driven by idiosyncratic visual similarities between the images. These results suggest that the hypothesis that naive subjects might be matching some of the interiors and exteriors based on conceptual similarities is a reasonable one. Indeed, as we discuss below, we believe that this may explain some of the cross-decoding in RSC.

Finally, a third possibility is that cross-decoding in scene regions is based on shared midlevel or high-level visual features. These might include architectural motifs and styles (Choo et al., 2015) and/or the shapes or textures of building materials, which might not be captured by models such as GIST and HMAX but could allow subjects to guess at the correspondences among images. As discussed below, we believe that such midlevel and high-level features might explain response in OPA, where cross-decoding was observed in all three experiments independent of familiarity or task.

Discussion

The primary goal of this study was to identify a neural mechanism for landmark recognition in the human brain. We postulated that such a mechanism would exhibit three characteristics. First, because landmarks are defined by their stable relationship to a spatial location or heading, a landmark recognition mechanism should treat different stimuli associated with a specific place as representationally similar, even when they are perceptually distinct. Second, this generalization across stimuli should be based on experience: subjects must know (or have reason to hypothesize) that the stimuli correspond to the same landmark. Third, this generalization must reflect a true common code, rather than simply being the byproduct of mnemonic association or mental imagery. Our results indicate that the PPA exhibits all of these characteristics.

The PPA has been previously implicated in landmark identification based on the fact that it responds strongly to scenes and buildings (Aguirre et al., 1998; Epstein and Kanwisher, 1998) and also to objects that would be suitable as landmarks (Janzen and Van Turenout 2004; Troiani et al., 2012). However, the current study provides critical new evidence for the role of the PPA in landmark identification by demonstrating for the first time that the PPA generalizes across perceptually dissimilar stimuli corresponding to the same landmark (specifically, the interior and exterior views of the same building). This finding suggests that the PPA extracts a common identity code from these two dissimilar stimuli. Moreover, the fact that cross-decoding between interiors and exteriors in the PPA is affected by familiarity with the landmark further supports the idea that the PPA performs landmark identification because familiarity is necessary to understand the correspondences between the interiors and the exteriors. Although we focus here on buildings, such a mechanism might be useful for generalizing across any set of stimuli that correspond to a specific place in the world, including different views of a street, courtyard, or landscape.

How does this abstract identity code in PPA, which reflects high-level knowledge about landmarks and scenes, fit with previous observations that PPA represents visual properties such as

retinotopic position or specific visual features? Most notably, our results suggest that PPA's responses are not determined exclusively by these visual properties. Instead, we suggest that abstract coding in the PPA complements visual representations of the appearance of landmarks, scene statistics, and scene layout within the same region (Epstein et al., 2003; Walther et al., 2009; Kravitz et al., 2011; Park et al., 2011; Rajimehr et al., 2011; Cant and Xu, 2012; Nasr et al., 2014), thus allowing a seamless transition from perceptual to conceptual or spatial content during landmark recognition. For example, the PPA's bias toward processing scene features in the upper visual field (Arcaro et al., 2009; Silson et al., 2015) might facilitate landmark recognition because landmarks typically appear at a distance and along the horizon. Notably, cross-decoding in our experiment was strongest in the anterior part of the PPA located within parahippocampal cortex proper, suggesting that this region might be more involved in coding abstract identity, in contrast to the more posterior portion of the PPA, which might be more important for coding the perceptual appearances of landmarks. This division is consistent with previous work showing that anterior PPA activates during the processing of abstract or spatial qualities of a stimulus (Bar and Aminoff, 2003; Davachi et al., 2003; Janzen and van Turenout, 2004; Aminoff et al., 2007; Fairhall et al., 2013), whereas posterior PPA activates during processing of visual qualities (Arcaro et al., 2009; Rajimehr et al., 2011; Cant and Xu, 2012; Nasr et al., 2014), and also with observations that anterior and posterior PPA can be distinguished by their differential functional connectivity to memory and visual processing networks (Baldassano et al., 2013; Nasr et al., 2013).

An unresolved question is the nature of the experience necessary for the PPA to form an identity code for a landmark. At first glance, the fact that cross-decoding was significant in Penn students but not in Temple students suggests that real-world experience with the landmark is necessary. However, this conclusion must be qualified by the fact that some degree of cross-decoding was observed in Temple students in the second half of Experiment 2, when these subjects were viewing the landmark exteriors and interiors for the second time. This suggests that visual experience alone, even in the absence of real-world navigation, might suffice to allow some degree of landmark generalization in the PPA. Indeed, previous work has implicated the PPA/PHC in rapid learning of associations between initially unfamiliar scenes (Turk-Browne et al., 2012). Although we cannot resolve this issue here, one possibility is that landmark representations in Penn students reflect long-term knowledge about landmark identity, whereas landmark representations in Temple students reflect top-down hypotheses about which scenes correspond to the same landmark—hypotheses that might direct on-the-fly attention toward perceptual features common to the interior and exterior of each building (Peelen et al., 2009; Çukur et al., 2013).

Two other scene regions, RSC and OPA, also showed evidence for landmark generalization, but the pattern across experiments was different from that observed in the PPA. Cross-decoding in RSC was not affected by personal experience with the landmarks but was affected by the performance of a concomitant memory retrieval task. Indeed, cross-decoding in RSC was abolished in Penn students in Experiment 3 (at least initially) when they had to retrieve faces associated with the landmarks. This suggests that rather than representing the landmark itself, RSC may represent the mnemonic context associated with the landmark (Maguire et al., 1999; Bar, 2007; Vann et al., 2009; Ranganath and Ritchey, 2012; Aminoff, 2014). Activation of this mnemonic con-

text is not obligatory but requires an additional act of memory retrieval that is susceptible to cognitive interruption.

A salient mnemonic context for a familiar landmark is knowledge about the broader spatial world surrounding it, which is not depicted in the visual stimulus but learned through navigational experience. Consistent with this view, increased response in RSC has been observed when subjects explicitly retrieve spatial information that allows them to orient themselves within a remembered or imagined spatial environment (Wolbers and Büchel, 2005; Spiers and Maguire, 2006; Byrne et al., 2007; Epstein et al., 2007; Hassabis et al., 2007; Epstein, 2008). Moreover, recent studies have found that RSC represents spatial quantities such as position or heading that only have meaning when defined relative to an extended spatial frame (Baumann and Mattingley, 2010; Vass and Epstein, 2013; Marchette et al., 2014) and responds especially strongly to permanent landmarks that might anchor such a frame (Auger et al., 2012). In the current experiment, however, the mnemonic context coded by RSC is unlikely to be the spatial coordinates of the stimuli because, in contrast to most of these previous studies, our subjects were not explicitly required to retrieve this information. Instead, we conjecture that our subjects used a conceptual rather than a spatial code to contextualize the stimuli (Fairhall and Caramazza, 2013; Fairhall et al., 2013; Aminoff, 2014). Consistent with this idea, naive subjects judged that interiors and exteriors corresponding to the same building depicted similar categories of place. In this view, RSC represents a semantic “space” in which the objects or actions associated with that category of place (e.g., bookstore) are encoded (Bar, 2007; Binder et al., 2009; Ranganath and Ritchey, 2012; Aminoff, 2014), and cross-decoding is possible because the interior and exterior of a building elicited similar semantic associations. In any case, our results suggest that RSC plays a very different role from PPA in landmark processing.

Cross-decoding in OPA was observed consistently across all three experiments, unaffected by familiarity and task. This pattern of results suggests that OPA processes mid-level perceptual features common to the interior and exterior scenes. Although we chose the interiors and exteriors to be as visually dissimilar as such stimuli typically are in the real world, and low-level visual models could not cross-decode the images, close inspection of the images reveals that, in some cases, there are common features (building materials, windows, and architectural motifs) that might be leveraged for generalization. Such features might be implicit though not linearly decodable in EVC and transformed into an explicit form in OPA (DiCarlo and Cox, 2007). Such a purely perceptual mechanism would be unaffected by high-level knowledge about which scenes correspond to the same place or by the mnemonic demands of the task and may perform visual analyses useful for scene recognition more generally. Support for this proposition comes from previous work indicating that OPA codes features characteristic of scenes (Kravitz et al., 2011; Bettencourt and Xu, 2013; Dilks et al., 2013; Ganaden et al., 2013; Choo et al., 2015).

In summary, our results reveal a tripartite division of labor whereby the PPA supports a landmark identity code that represents objects or topographical elements that signify a particular place, RSC retrieves spatial and conceptual information about these places, and OPA represents their perceptual details. These findings clarify how we represent the landmarks that mark the distinct locations we encounter in our daily lives.

References

- Aguirre GK (2007) Continuous carry-over designs for fMRI. *Neuroimage* 35:1480–1494. [CrossRef Medline](#)
- Aguirre GK, D'Esposito M (1999) Topographical disorientation: a synthesis and taxonomy. *Brain* 122:1613–1628. [CrossRef Medline](#)
- Aguirre GK, Zarahn E, D'Esposito M (1998) An area within human ventral cortex sensitive to “building” stimuli: evidence and implications. *Neuron* 21:373–383. [CrossRef Medline](#)
- Aminoff E, Gronau N, Bar M (2007) The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb Cortex* 17:1493–1503. [CrossRef Medline](#)
- Aminoff EM (2014) Putting scenes in context. In: *Scene vision: making sense of what we see* (Kestutis K, Bar M, eds.), p 135. Cambridge, MA: MIT.
- Amit E, Mehoudar E, Trope Y, Yovel G (2012) Do object-category selective regions in the ventral visual stream represent perceived distance information? *Brain Cogn* 80:201–213. [CrossRef Medline](#)
- Arcaro MJ, McMains SA, Singer BD, Kastner S (2009) Retinotopic organization of human ventral visual cortex. *J Neurosci* 29:10638–10652. [CrossRef Medline](#)
- Auger SD, Maguire EA (2013) Assessing the mechanism of response in the retrosplenial cortex of good and poor navigators. *Cortex* 49:2904–2913. [CrossRef Medline](#)
- Auger SD, Mullally SL, Maguire EA (2012) Retrosplenial cortex codes for permanent landmarks. *PLoS One* 7:e43620. [CrossRef Medline](#)
- Baldassano C, Beck DM, Fei-Fei L (2013) Differential connectivity within the parahippocampal place area. *Neuroimage* 75:228–237. [CrossRef Medline](#)
- Bar M (2007) The proactive brain: using analogies and associations to generate predictions. *Trends Cogn Sci* 11:280–289. [CrossRef Medline](#)
- Bar M, Aminoff E (2003) Cortical analysis of visual context. *Neuron* 38:347–358. [CrossRef Medline](#)
- Baumann O, Mattingley JB (2010) Medial parietal cortex encodes perceived heading direction in humans. *J Neurosci* 30:12897–12901. [CrossRef Medline](#)
- Bettencourt KC, Xu Y (2013) The role of transverse occipital sulcus in scene perception and its relationship to object individuation in inferior intraparietal sulcus. *J Cogn Neurosci* 25:1711–1722. [CrossRef Medline](#)
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796. [CrossRef Medline](#)
- Byrne P, Becker S, Burgess N (2007) Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychol Rev* 114:340–375. [CrossRef Medline](#)
- Cant JS, Xu Y (2012) Object ensemble processing in human anterior-medial ventral visual cortex. *J Neurosci* 32:7685–7700. [CrossRef Medline](#)
- Cate AD, Goodale MA, Köhler S (2011) The role of apparent size in building- and object-specific regions of ventral visual cortex. *Brain Res* 1388:109–122. [CrossRef Medline](#)
- Choo H, Nikrahe B, Nasar J, Walther D (2015) Neural decoding of architectural styles from scene-specific brain regions. *J Vis* 15:520. [CrossRef Medline](#)
- Çukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci* 16:763–770. [CrossRef Medline](#)
- Davachi L, Mitchell JP, Wagner AD (2003) Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc Natl Acad Sci U S A* 100:2157–2162. [CrossRef Medline](#)
- DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. *Trends Cogn Sci* 11:333–341. [CrossRef Medline](#)
- Dilks DD, Julian JB, Paunov AM, Kanwisher N (2013) The occipital place area is causally and selectively involved in scene perception. *J Neurosci* 33:1331–1336a. [CrossRef Medline](#)
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601. [CrossRef Medline](#)
- Epstein R, Graham KS, Downing PE (2003) Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* 37:865–876. [CrossRef Medline](#)
- Epstein RA (2008) Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci* 12:388–396. [CrossRef Medline](#)
- Epstein RA, Morgan LK (2012) Neural responses to visual scenes reveals

- inconsistencies between fMRI adaptation and multivoxel pattern analysis. *Neuropsychologia* 50:530–543. [CrossRef Medline](#)
- Epstein RA, Vass LK (2014) Neural systems for landmark-based wayfinding in humans. *Philos Trans R Soc Lond B Biol Sci* 369:20120533. [CrossRef Medline](#)
- Epstein RA, Parker WE, Feiler AM (2007) Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *J Neurosci* 27:6141–6149. [CrossRef Medline](#)
- Fairhall SL, Caramazza A (2013) Brain regions that represent amodal conceptual knowledge. *J Neurosci* 33:10552–10558. [CrossRef Medline](#)
- Fairhall SL, Anzellotti S, Ubaldi S, Caramazza A (2013) Person- and place-selective neural substrates for entity-specific semantic access. *Cereb Cortex* 24:1687–1696. [CrossRef Medline](#)
- Gallistel CR (1990) *The organization of learning*. Cambridge, MA: MIT.
- Ganaden RE, Mullin CR, Steeves JK (2013) Transcranial magnetic stimulation to the transverse occipital sulcus affects scene but not object processing. *J Cogn Neurosci* 25:961–968. [CrossRef Medline](#)
- Hassabis D, Kumaran D, Maguire EA (2007) Using imagination to understand the neural basis of episodic memory. *J Neurosci* 27:14365–14374. [CrossRef Medline](#)
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430. [CrossRef Medline](#)
- Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkänen A (1998) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *Am J Neuroradiol* 19:659–671. [Medline](#)
- Janzen G, van Turennout M (2004) Selective neural representation of objects relevant for navigation. *Nat Neurosci* 7:673–677. [CrossRef Medline](#)
- Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–841. [CrossRef Medline](#)
- Julian JB, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60:2357–2364. [CrossRef Medline](#)
- Konkle T, Oliva A (2012) A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74:1114–1124. [CrossRef Medline](#)
- Kravitz DJ, Peng CS, Baker CI (2011) Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *J Neurosci* 31:7322–7333. [CrossRef Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef Medline](#)
- Lynch K (1960) *The image of the city*. Cambridge, MA: MIT.
- Maguire EA, Frith CD, Morris RG (1999) The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. *Brain* 122:1839–1850. [CrossRef Medline](#)
- Marchette SA, Vass LK, Ryan J, Epstein RA (2014) Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nat Neurosci* 17:1598–1606. [CrossRef Medline](#)
- Morgan LK, Macevoy SP, Aguirre GK, Epstein RA (2011) Distances between real-world locations are represented in the human hippocampus. *J Neurosci* 31:1238–1245. [CrossRef Medline](#)
- Mullally SL, Maguire EA (2011) A new role for the parahippocampal cortex in representing space. *J Neurosci* 31:7441–7449. [CrossRef Medline](#)
- Nasr S, Devaney KJ, Tootell RB (2013) Spatial encoding and underlying circuitry in scene-selective cortex. *Neuroimage* 83:892–900. [CrossRef Medline](#)
- Nasr S, Echarvarria CE, Tootell RB (2014) Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *J Neurosci* 34:6721–6735. [CrossRef Medline](#)
- Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping* 15:1–25. [CrossRef Medline](#)
- Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int J Comput Vis* 42:145–175. [CrossRef](#)
- Park S, Brady TF, Greene MR, Oliva A (2011) Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *J Neurosci* 31:1333–1340. [CrossRef Medline](#)
- Peelen MV, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460:94–97. [CrossRef Medline](#)
- Pruessner JC, Köhler S, Crane J, Pruessner M, Lord C, Byrne A, Kabani N, Collins DL, Evans AC (2002) Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. *Cereb Cortex* 12:1342–1353. [CrossRef Medline](#)
- Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RB (2011) The “parahippocampal place area” responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol* 9:e1000608. [CrossRef Medline](#)
- Ranganath C, Ritchey M (2012) Two cortical systems for memory-guided behaviour. *Nat Rev Neurosci* 13:713–726. [CrossRef Medline](#)
- Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2:1019–1025. [CrossRef Medline](#)
- Schinazi VR, Epstein RA (2010) Neural correlates of real-world route learning. *Neuroimage* 53:725–735. [CrossRef Medline](#)
- Siegel AW, White SH (1975) The development of spatial representations of large-scale environments. *Adv Child Dev Behav* 10:9–55. [CrossRef Medline](#)
- Silson EH, Chan AW, Reynolds RC, Kravitz DJ, Baker CI (2015) A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *J Neurosci* 35:11921–11935. [CrossRef Medline](#)
- Spiers HJ, Maguire EA (2006) Thoughts, behaviour, and brain dynamics during navigation in the real world. *Neuroimage* 31:1826–1840. [CrossRef Medline](#)
- Theriault C, Thome N, Cord M (2011) HMAX-S: deep scale representation for biologically inspired image categorization. Paper presented at Image processing (ICIP), 18th IEEE International Conference, Brussels, Belgium, September.
- Troiani V, Stigliani A, Smith ME, Epstein RA (2012) Multiple object properties drive scene-selective regions. *Cereb Cortex* 24:883–897. [CrossRef Medline](#)
- Turk-Browne NB, Simon MG, Sederberg PB (2012) Scene representations in parahippocampal cortex depend on temporal context. *J Neurosci* 32:7202–7207. [CrossRef Medline](#)
- Van Leemput K, Bakkour A, Benner T, Wiggins G, Wald LL, Augustinack J, Dickerson BC, Golland P, Fischl B (2009) Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* 19:549–557. [CrossRef Medline](#)
- Vann SD, Aggleton JP, Maguire EA (2009) What does the retrosplenial cortex do? *Nat Rev Neurosci* 10:792–802. [CrossRef Medline](#)
- Vass LK, Epstein RA (2013) Abstract representations of location and facing direction in the human brain. *J Neurosci* 33:6133–6142. [CrossRef Medline](#)
- Walther DB, Caddigan E, Fei-Fei L, Beck DM (2009) Natural scene categories revealed in distributed patterns of activity in the human brain. *J Neurosci* 29:10573–10581. [CrossRef Medline](#)
- Wolbers T, Büchel C (2005) Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *J Neurosci* 25:3333–3340. [CrossRef Medline](#)