Coding of navigational affordances in the human visual system

Michael F. Bonner[a,1] and Russell A. Epstein[a]

[a]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104

A central component of spatial navigation is determining where one can and cannot go in the immediate environment. We used fMRI to test the hypothesis that the human visual system solves this problem by automatically identifying the navigational affordances of the local scene. Multivoxel pattern analyses showed that a scene-selective region of dorsal occipitoparietal cortex, known as the occipital place area, represents pathways for movement in scenes in a manner that is tolerant to variability in other visual features. These effects were found in two experiments: One using tightly controlled artificial environments as stimuli, the other using a diverse set of complex, natural scenes. A reconstruction analysis demonstrated that the population codes of the occipital place area could be used to predict the affordances of novel scenes. Taken together, these results reveal a previously unknown mechanism for perceiving the affordance structure of navigable space.

scene-selective visual cortex | occipital place area | affordances | navigation | dorsal stream

It has long been hypothesized that perceptual systems are optimized for the processing of features that afford ecologically important behaviors (1, 2). This perspective has gained renewed support from recent work on action planning, which suggests that the action system continually prepares multiple, parallel plans that are appropriate for the environment (3, 4). If this view is correct, then sensory systems should be routinely engaged in identifying the potential of the environment for action, and they should explicitly and automatically encode these action affordances. Here we explore this idea for spatial navigation, a behavior that is essential for survival and ubiquitous among mobile organisms.

A critical component of spatial navigation is the ability to understand where one can and cannot go in the local environment: for example, knowing that one can exit a room through a corridor or a doorway but not through a window or a painting on the wall. We reasoned that if perceptual systems routinely extract the parameters of the environment that delimit potential actions, then these navigational affordances should be automatically encoded during scene perception, even when subjects are not engaged in a navigational task.

Previous work has shown that observers can determine the overall navigability of a scene—for example, whether it is possible to move through the scene or not—from a brief glance (5). However, no study has examined the coding of fine-grained navigational affordances, such as whether the direction one can move in the scene is to the left or to the right. Furthermore, only recently have investigators begun to characterize the affordance properties of scenes, and this work has focused not on navigational affordances but on more abstract behavioral events that can be used to define scene categories, such as cooking and sleeping (6). It therefore remains unknown whether navigational affordances play a fundamental role in the perception of visual scenes, and if so, which brain systems are involved.

To address this issue, we used multivoxel pattern analysis of fMRI data in two experiments. In Exp. 1, subjects were scanned while viewing artificially rendered images of rooms that varied in the locations of the visible exits. In Exp. 2, subjects were scanned while viewing natural photographs of rooms that varied in regard to the paths that one could take to walk through the space. In both experiments, subjects performed tasks that made no reference to these affordances and were unrelated to navigation. Despite the orthogonal tasks, we predicted that the navigational affordances of the scenes would be automatically extracted and explicitly coded in the visual system. To test this hypothesis, we attempted to identify fMRI activation patterns that distinguished between scenes based on the spatial structure of their navigational affordances, but generalized over other scene properties. To anticipate, our data suggest that these representations are indeed automatically extracted, and that the strongest representation of navigational affordances is in a scene-selective region of visual cortex known as the occipital place area (OPA).

Results

Navigational Affordances in Artificially Rendered Environments. The goal of the first experiment was to identify representations of navigational affordances in the visual system. To do this, we created a set of tightly controlled, artificially rendered stimuli that allowed us to systematically manipulate navigational affordances while maintaining complete control over other visual properties of the scenes. Specifically, we designed 3D models of virtual environments and generated high-resolution renderings of these environments with a variety of surface textures (Fig. 1). All environments had the same coarse local geometry as defined by the spatial layout of the walls, but the locations and number of the visible exits were varied to create eight stimulus conditions that differed in their navigational affordance structure. To ensure that differences between these conditions could not be trivially explained by differences in the spatial distribution of low-level visual features across

Significance

As we move about the world, we use vision to determine where we can go in our local environment and where our path is blocked. Despite the ubiquity of this problem, little is known about how it is solved by the brain. Here we show that a specific region in the human visual system, known as the occipital place area, automatically encodes the structure of navigable space in visual scenes, thus providing evidence for a bottom-up visual mechanism for perceiving potential paths for movement in one's immediate surroundings. These findings are consistent with classic theoretical work predicting that sensory systems are optimized for processing environmental features that afford ecologically relevant behaviors, including the fundamental behavior of navigation.
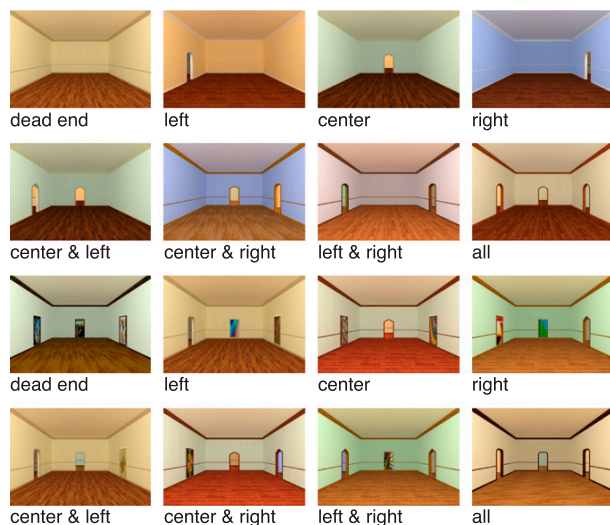
NEUROSCIENCE

Fig. 1. Examples of artificially rendered environments used as stimuli in Exp. 1. Eight navigational-affordance conditions were defined by the number and position of open doorways along the walls. For each condition, we created 18 aesthetic variants that differed in surface textures and the shapes of the doorways (one shown for each condition), and for each of these aesthetic variants, we created one stimulus in which walls with no exit were blank (*Top* two rows) and one stimulus in which walls with no exit contained an abstract painting (*Bottom* two rows).

the image, half of the stimuli in each condition included paintings in locations not occupied by doorways. These paintings were approximately the same size and shape as the doors, but did not afford egress from the room.

Twelve subjects (eight female) viewed images of these environments while being scanned on a Siemens 3.0 T Trio MRI scanner using a protocol sensitive to blood oxygenation level-dependent (BOLD) contrasts (see *SI Methods* for details). All participants provided written informed consent in compliance with procedures approved by the University of Pennsylvania Institutional Review Board. Stimuli were presented for 2 s each while subjects maintained central fixation and performed an orthogonal task in which they indicated whether two dots that appeared overlaid on the scene were the same color (Fig. S1). The dots appeared in varying locations, requiring subjects to distribute their attention across the visual field.

We focused our initial analyses on three regions that have previously been shown to be strongly involved in scene processing: the OPA, the parahippocampal place area (PPA), and the retrosplenial complex (RSC) (7–11). All three of these regions respond more strongly to the viewing of spatial scenes than other stimuli, such as objects and faces, and are thus good candidates for supporting representations of navigational affordances. As a control, we also examined activity patterns in early visual cortex (EVC). Scene regions were identified based on their greater response to scenes than objects in independent localizer scans, and the EVC was defined based on its greater response to scrambled objects than scenes. To probe the information content of these regions of interest (ROIs) we created representational dissimilarity matrices (RDMs) through pairwise comparisons of neural activity patterns for each condition, and then calculated the correlation of these RDMs with a model that quantified the similarity of navigational affordances based on the number and locations of the rooms' exits (Fig. 2*A* and *SI Methods*). (See *Whole-Brain Searchlight Analyses* below for results outside of these ROIs.)

This analysis revealed strong evidence for the coding of navigational affordances in the OPA (Fig. 2*B*) [$t(11) = 8.06$, $P =$

0.000003], but no reliable effects in the PPA, RSC, or EVC (all $P > 0.24$). Direct comparisons showed that this effect was significantly stronger in the OPA than in each of the other ROIs (all $P < 0.026$). The neural RDMs in this analysis were constructed from comparisons of stimulus sets that differed in surface textures and, in some cases, also differed in the identity or presence of paintings along the walls (*SI Methods*). In follow-up analyses, we found that the OPA was the only ROI to exhibit significant effects for the coding of navigational affordances across all manipulations of textures and paintings (Fig. S2). The scenes with paintings are particularly informative because they control for low-level visual differences across affordance conditions. When analyses were restricted to comparisons between scenes with paintings, the mean representational similarity analysis (RSA) effect in the OPA was similar to that in Fig. 2 (Fig. S2*C*). When analyses were restricted to comparisons between scenes with paintings and scenes without paintings, the mean RSA effect in the OPA was reduced by half (Fig. S2*D*).

We also examined univariate responses and found that, for scenes without paintings, the mean response in most ROIs increased in relation to the number of doorways, but this was not observed in any ROI for scenes with paintings, which are better matched on visual complexity (Fig. S3). Thus, fine-grain affordance information can be detected in multivariate pattern analyses, but it is not a prominent component of the average univariate response in an ROI.

Together, these results demonstrate that the OPA extracts spatial features from visual scenes that can be used to identify the navigational affordances of the local environment. These representations appear to reflect fine-grained information about the structure of navigable space, allowing the OPA to distinguish between scenes that have the same coarse geometric shape but differ on the layout of visible exits. Furthermore, these findings demonstrate that affordance representations in the OPA exhibit some degree of tolerance to the other visual properties of scenes, suggesting a general-purpose mechanism for coding the navigational structure of space across a range of contexts.

We were surprised that we did not observe reliable evidence for navigational-affordance coding in the PPA, which has long been hypothesized to represent the spatial layout of scenes (8, 12–14). One possibility is that the PPA is sensitive to the coarse shape of the local environment as defined by the walls, but is insensitive to fine-grained manipulations of spatial structure as defined by the locations of exits. Another (nonexclusive) possibility is that the PPA encodes higher-level, conjunctive representations of the textural and structural features of scenes (15, 16), making it unsuited for generalizing across scenes with the same affordance structure but different textures (which might be perceived as being different places). This finding would be consistent with the hypothesis that the PPA supports the identification of familiar places
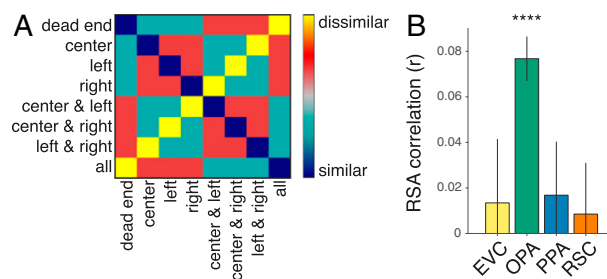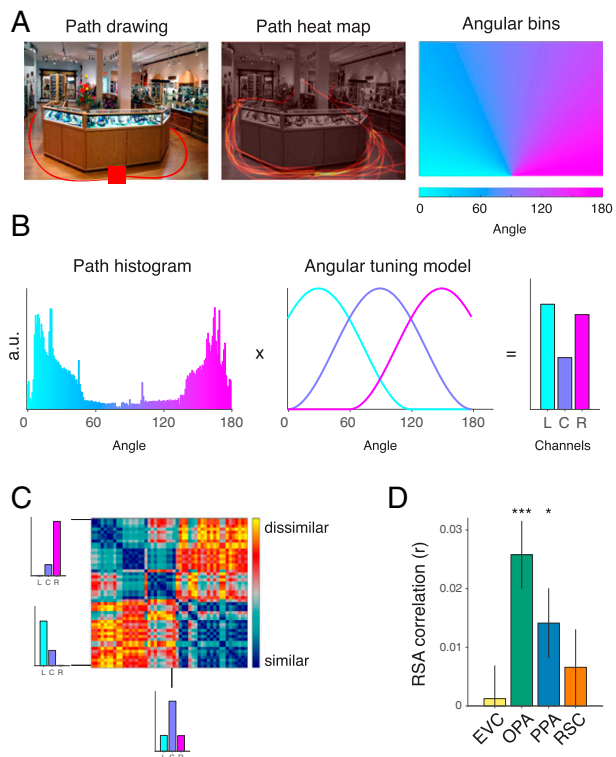

Fig. 2. Coding of navigational affordances in artificially rendered environments. (*A*) Model RDM of navigational affordances defined by overlap in the locations of the open doorways. (*B*) RSA of this model RDM in each ROI. The OPA showed a strong and reliable effect for the coding of navigational affordances. Error bars represent ± 1 SEM; ****$P < 0.0001$.

and landmarks, as both textural and structural features contribute to place identity (17). To address the latter possibility, we performed a follow-up analysis of neural dissimilarities in the PPA and OPA to examine the relationship between texture and navigational affordance coding in these regions (*SI Methods*). This analysis suggested that the PPA was sensitive to the conjunction of textural and structural scene features, as evidenced by a trend toward an interaction between these factors, and the fact that affordance coding was only found for scenes of the same texture. In contrast, the OPA encodes affordance structure regardless of textural differences (Fig. S4). This finding explains the lack of reliable effects in the PPA for the original set of analyses, and it suggests that the PPA may use a different mechanism for scene processing than the OPA. Specifically, the PPA may perform a combinatorial analysis of the multiple scene features that identify a place or landmark.

Navigational Affordances in Complex, Natural Images. The artificial stimuli used in Exp. 1 allowed for a tightly controlled design with specific manipulations of scene properties. This was useful for the initial identification of scene affordance representations. However, an important question is whether the results would generalize to more complex, naturalistic scenes. We addressed this in the second experiment by testing for the coding of navigational affordances using photographs of real-world environments. Specifically, we examined fMRI responses to 50 images of indoor scenes viewed at eye level with clear navigational paths emanating from the observer's point of view (Fig. S5). Sixteen new subjects (eight female) were scanned on a Siemens 3.0 T Prisma scanner while viewing these images for 1.5 s each. Subjects were asked to maintain central fixation while performing an orthogonal category-detection task (i.e., pressing a button whenever a scene was a bathroom) (Fig. S6).

An intuitive way to conceptualize the navigational affordances of real-world scenes is as a set of potential paths radiating along angles from the implied position of the viewer (1). For each image in Exp. 2, we asked a group of independent raters to indicate with a computer mouse the paths that they would take to walk through each environment starting from the bottom of the scene (Fig. 3A). From these responses we created heat maps that provided a statistical summary of the navigational trajectories in each image, and we then quantified these data over a range of angular directions radiating from the starting point at the bottom center of the image (Fig. 3B). We modeled affordance coding using as a set of explicitly defined encoding channels with tuning curves over the range of angular directions. To do this, we applied an approach that has previously been used to characterize the neural coding of color and orientation (18, 19). Specifically, we modeled navigational affordances using a basic set of response channels with tuning curves that were maximally sensitive to trajectories going to the left, center, or right (Fig. 3B and *SI Methods*). Scenes were then represented by the degree to which they drove the responses of these channels. This encoding model thus reduces the high-dimensional trajectory data to a small set of well-defined feature channels. We then created a model RDM based on pairwise comparisons of all scenes' affordance representations (Fig. 3C; see Fig. S7 for a visualization). The scenes used in the fMRI experiment were selected from a larger set of such labeled images to ensure that this navigational RDM was uncorrelated with RDMs derived from several models of low-level vision and spatial attention (*SI Methods*).

We used RSA to compare the navigational model RDM to the same set of visual ROIs examined in Exp. 1. This analysis showed a strong effect for the coding of navigational affordances in the OPA (Fig. 3 C and D) [$t(15) = 4.54$, $P = 0.0002$], replicating the main finding of the first experiment. Interestingly, we also observed a weaker but significant effect in the PPA [$t(15) = 2.42$, $P = 0.0144$], which may reflect the fact that the affordances in the

Fig. 3. Coding of navigational affordances in natural scenes. (*A*) In a norming study, a group of independent raters were asked to indicate the paths that they would take to walk through each scene starting from the bottom center of the image (*Left*). These data were combined across raters to produce a heat map of the possible navigational trajectories through each scene (*Center*). Angular histograms were created by summing the responses within a set of angular bins radiating from the bottom center of each image (*Right*). (*B*) The resulting histograms summarize the trajectory responses across the entire range of angles that are visible in the image. Navigational affordances were modeled using a set of hypothesized encoding channels with tuning curves that broadly code for trajectories to the left (L), center (C), and right (R). Model representations were computed as the product of the angular-histogram vectors and the tuning curves of the navigational-affordance channels. a.u., arbitrary units. (*C*) A model RDM was created by comparing the affordance-channel representations across images. (*D*) RSA showed that the strongest effect for the coding of navigational affordances was in the OPA. Error bars represent ± 1 SEM; *$P < 0.05$, ***$P < 0.001$.

real-world stimuli were not regulated to be independent of coarse-scale geometry or scene texture. Nonetheless, navigational affordances were encoded most reliably in the OPA, and direct comparisons show that affordance coding in the OPA was stronger than in each of the other ROIs (all $P < 0.042$).

We next performed a follow-up analysis to compare the coding of navigational affordances with representations from the Gist computer-vision model (20). The Gist model captures low-level image features that often covary with higher-level scene properties, such as semantic category and spatial expanse. Using partial-correlation analyses, we identified the RSA correlations that could be uniquely attributed to each model (*SI Methods*). The Gist model had a higher RSA correlation than the affordance model across all ROIs, with the largest effect in the EVC (Fig. S8A) [$t(15) = 7.68$, $P = 0.00000071$] and significant effects in the OPA [$t(15) = 3.15$, $P = 0.0033$] and PPA [$t(15) = 5.08$, $P = 0.000067$]. Critically, the effects of the affordance model remained significant in the OPA [$t(15) = 3.75$, $P = 0.00096$] and PPA [$t(15) = 1.82$, $P = 0.044$] when the variance of the Gist features was

NEUROSCIENCE

partialled out (Fig. S8A). Moreover, when the variance explained by the EVC was partialled out from the scene-selective ROIs (*SI Methods*), the effect of the affordance model remained significant in the OPA [$t(15) = 3.95$, $P = 0.00065$] and PPA [$t(15) = 3.10$, $P = 0.0036$], but the Gist model was no longer significant in either of these regions (Fig. S8B) (both $P > 0.35$). These results suggest that the features of the Gist model are coded in the EVC and preserved in downstream scene-selective regions, whereas navigational affordance features are extracted by computations within scene-selective cortices.

These findings show once again that the OPA encodes the navigational affordances of visual scenes by demonstrating that the effect observed with artificial stimuli generalizes to natural scene images with heterogeneous visual and semantic properties. The results also provide further support for the idea that this information is extracted automatically, even when subjects are not engaged in a navigational task (although the apparent automaticity of these effects does not rule out the possibility that affordance representations could be modulated by the demands of a navigational task). Finally, these findings demonstrate the feasibility of modeling the neural coding of navigational affordances in natural scenes using a simple set of response channels tuned to the egocentric angles of potential navigational trajectories.

**Reconstructing Navigational Affordances from Cortical Responses.** The findings presented above indicate that the population codes of scene-selective visual cortex contain fine-grained information about where one can navigate in visual scenes. This finding suggests that by using the multivoxel activation patterns of scene-selective ROIs, we should be able to reconstruct the navigational affordances of novel scenes regardless of their semantic content and other navigationally irrelevant visual properties. In other words, it should be possible to train a linear decoder that can predict the affordances of a previously unseen image. We tested this possibility by attempting to generate heat maps of the navigational affordances

of novel scenes from the activation patterns within an ROI. To maximize the prediction accuracy of this model, we first concatenated the ROI data across all subjects to create a multisubject response matrix, which contains a single activation vector for each training image that was created by combining the response patterns to that image across all subjects (Fig. 4A). We then performed a series of pixel-wise reconstruction analyses using principal component regression (*SI Methods*). We assessed prediction accuracy through a leave-one-stimulus-out (LOO) cross-validation procedure. We found that the responses of the OPA were sufficient to generate affordance reconstructions at an accuracy level that was well above chance, more so than the other scene regions and EVC (Fig. 4 *B* and *C*) ($P < 0.05$ permutation test). This finding suggests that the population codes of the OPA contain rich information about the affordance structure of the navigational environment, and that these representations could be used to map out potential trajectories in local space. As in the RSA findings above, the PPA also showed a weaker but significant effect for the reconstruction analysis ($P < 0.05$ permutation test), providing further evidence that the PPA may also contribute to affordance coding in natural scenes.

Whole-Brain Searchlight Analyses. The fMRI analyses above focus on the responses in scene-selective ROIs. To test for possible effects of navigational-affordance coding outside of our a priori ROIs, we also performed whole-brain searchlight analyses on the data from Exps. 1 and 2 (*SI Methods*). Searchlights were spheres with a 6-mm radius around each voxel; data were corrected for multiple comparisons across the entire brain using a permutation test to establish the true family-wise error (FWE) rate. In both experiments, we found evidence for the coding of navigational affordances in a region located at the junction of the intraparietal and transverse occipital sulci, corresponding to the dorsomedial boundary of the OPA (Fig. 5). No other regions showed significant effects in either experiment (see *SI Searchlight Results* and Fig. S9 for a more detailed examination of the searchlight clusters).
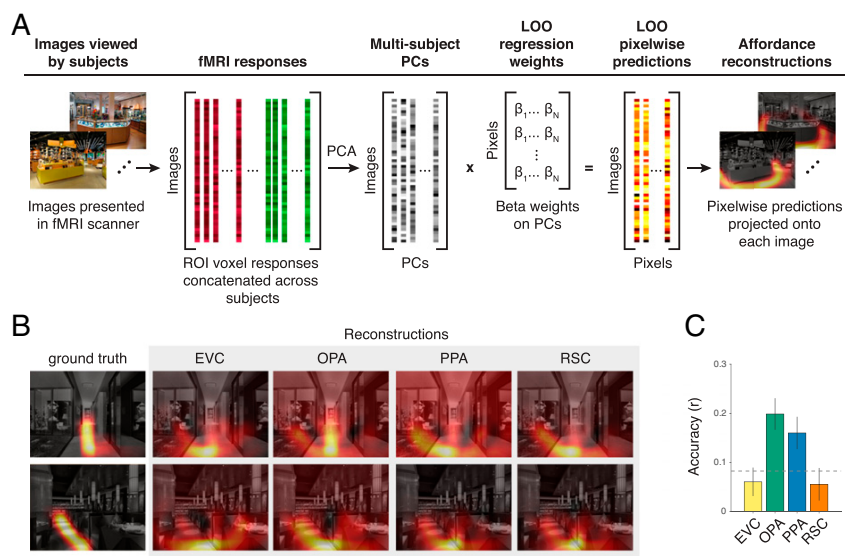
Fig. 4. Reconstruction of navigational-affordance maps. (*A*) Navigational-affordance maps were reconstructed from the fMRI responses within each ROI. First, a data-fusion procedure was used to create a set of multisubject ROI responses for use in the reconstruction model. For each ROI, principal component analysis (PCA) was applied to a matrix of voxel responses concatenated across all subjects. The resulting multisubject PCs were used as predictors in a set of pixelwise decoding models. These decoding models used linear regression to generate intensity values for individual pixels from a weighted sum of multi-subject fMRI responses. The ability of the models to reconstruct the affordances of novel stimuli was assessed through LOO cross-validation on each image in turn. (*B*) Example reconstructions of navigational-affordance maps from the cortical responses in each ROI. (*C*) Affordance maps were reconstructed most accurately from the responses of the OPA. Bars represent the mean accuracy across images. Error bars represent ± 1 SEM. The dashed line indicates chance performance at $P < 0.05$ permutation-test FWE.
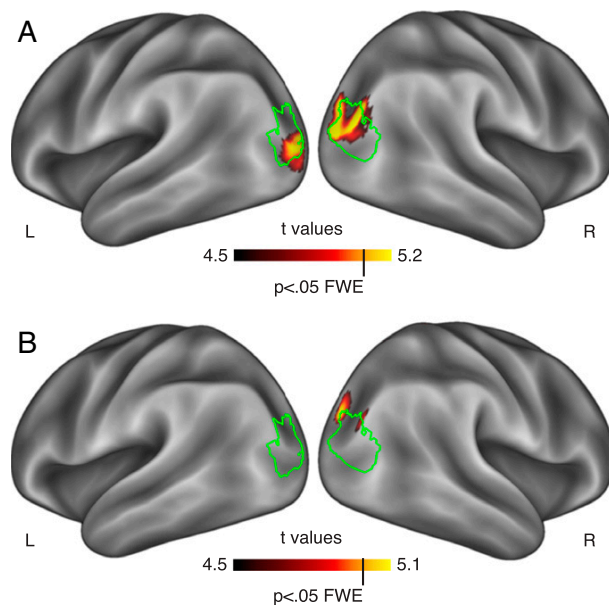
Fig. 5. Whole-brain searchlight analyses. (*A*) RSA of navigational-affordance coding for the artificially rendered environments of Exp. 1. (*B*) RSA of navigational-affordance coding for the natural scene images of Exp. 2. Consistent with the ROI findings, both experiments show whole-brain corrected effects at the junction of the intraparietal and transverse-occipital sulci. The green outlines correspond to the borders of the OPA parcels.

Discussion

The principal goal of this study was to identify representations of navigational affordances in the human visual system. We found evidence for such representations in the OPA, a region of scene-selective visual cortex located near the junction of the transverse occipital and intraparietal sulci. Scenes with similar navigational affordances (i.e., similar pathways for movement) elicited similar multivoxel activation patterns in the OPA, whereas scenes with different navigational affordances elicited multivoxel activation patterns that were more dissimilar. This effect was observed in Exp. 1 for virtual rooms that all had the same coarse geometry but differed in the locations of the exits, and it was replicated in Exp. 2 for real-world indoor environments that varied on multiple visual and semantic dimensions. Indeed, in Exp. 2, it was possible to reconstruct the affordances of novel environments using the multivoxel activation patterns of the OPA. Crucially, in both experiments, affordance representations were observed even though subjects performed tasks that did not require them to plan routes through the environment. This finding suggests that navigational affordances are encoded automatically even when they are not directly task-relevant. Together, these findings reveal a mechanism for automatically identifying the affordance structure of the local spatial environment.

Previous work on affordances has focused largely on the actions afforded by objects, not scenes. Behavioral studies have shown that visual objects prime grasp-related information that reflects their orientation and function (21), and neural evidence suggests that objects automatically activate the visuospatial and motor regions that mediate their commonly associated actions (22). Like objects, spatial scenes also afford a set of potential actions, and one of the most fundamental of these is navigation (5). In fact, some of the earliest work on the theory of affordances focused on aspects of the spatial environment that afford locomotion (1). However, little work has examined how the brain maps out the affordances of navigable space. Here we provide evidence for the

neural coding of such affordances, demonstrating that the human visual system routinely extracts the visuospatial information needed to map out the potential paths in a scene. This proposed mechanism also aligns well with a recent theory of action-planning known as the affordance-competition framework, which suggests that observers routinely encode multiple, parallel plans of the relevant actions afforded by their environment and then rapidly select among these when implementing a behavior (4).

The fact that navigational affordance representations were anatomically localized to the OPA has important implications for our understanding of the functional organization of the visual system. The OPA, along with the PPA and RSC, constitute a network of brain regions that show a strong response to visual scenes (7, 11) [note that in the literature, the OPA has also been referred to as the transverse occipital sulcus (10)]. The OPA is located in the dorsal visual stream, a cluster of processing pathways that has been broadly associated with visuospatial perception and visuomotor processing (23, 24). Many studies have examined the contribution of the dorsal stream to visually guided actions of the eyes, hands, and arms (25), but little is known about its possible role in extracting visuospatial parameters useful for navigation. It has recently been proposed that, within the scene-selective network, the OPA is particularly well suited for guiding navigation in local space (26), although there is also evidence that it plays a role in scene categorization (27, 28). The OPA is sensitive to changes in the chirality (i.e., mirror-image flips) and egocentric depth of visual scenes (26, 29), both of which reflect changes in navigationally relevant spatial parameters. A recent study using transcranial magnetic stimulation showed that the OPA has a causal role in perceiving spatial relationships between objects and environmental boundaries, but not between objects and other objects (30). Furthermore, the OPA is known to show a retinotopic bias for information in the periphery, which could facilitate the perception of extended spatial structures, and also a bias for the lower visual field, where paths tend to be (31). However, no previous study has investigated whether the OPA, or any other region, encodes the structural arrangement of the paths and boundaries in a scene. Our experiments show that the OPA does indeed encode these environmental features, thus providing critical information about where one can and cannot go in a visual scene.

In contrast, we observed weaker evidence for the coding of navigational affordances in the PPA and RSC. The null effect of affordance coding in RSC is consistent with previous work indicating that this region primarily supports spatial-memory retrieval rather than scene perception (32–34). The absence of robust affordance coding in the PPA, on the other hand, is at first glance surprising, as it has long been hypothesized that the PPA encodes the geometric layout of scenes (8, 35), possibly in the service of guiding navigation in local space (8, 12, 13, 36). And previous work indicates that the PPA represents at least some coarse aspects of scene geometry (12, 13). A possible interpretation of the current results is that the PPA supports affordance representations that are coarser than those of the OPA, and thus do not reflect the detailed affordance structure of scenes. Alternatively, the PPA may encode fine-grained geometry related to affordances, but these representations might not have been revealed in the current experiments because they do not generalize across scenes that are interpreted as being different places. Supporting the latter view, the secondary analysis of Exp. 1 suggests that the PPA supports combinatorial representations of multiple scene features, including both affordance structure and visual textures. Such combinatorial representations would be especially useful for place recognition, given that places are more likely to be distinguishable based on combinations of shape and texture rather than shape or texture alone. Indeed, recent work shows that the PPA contains representations of landmark identity, which may reflect knowledge of the many visual features that are associated with a familiar place (16, 17, 37, 38).

NEUROSCIENCE

The findings from our first experiment lend additional support to this theory.

Several key questions remain for future work. First, what is the coordinate frame used by the OPA to code navigational affordances? Regions within the dorsal visual stream are known to code the positions of objects and other environmental features relative to various reference points, including the eyes, head, hands, and world (25). Our data suggest that the OPA may encode affordances in an egocentric reference frame but this remains to be tested. Second, are the responses of the OPA best explained by a stable set of visual-feature preferences, or can these feature preferences be flexibly modulated by navigational demands? Third, what is the role of memory in the analysis of navigational affordances? Affordances in our stimuli could be identified solely from visual features, but in other situations spatial knowledge, such as memory for where a path leads, might come into play, as might semantic knowledge about material properties (e.g., the navigability of pavement, sand, or water). Memory systems might also be crucial for using affordance information to actively plan paths through the local environment to reach specific goals (39). Fourth, how do navigational affordances influence eye movements and attention? We attempted to minimize the contribution of these factors by asking subjects to maintain central fixation, and by examining images in which salient visual features were uncorrelated with navigational affordances, but in other circumstances, affordances and attention will likely interact. Fifth, what higher-level systems do these OPA representations feed into and how are these representations used? Within the dorsal stream, the OPA straddles several retinotopically defined regions, including V3A, V3B, LO1, LO2, and V7, and may also extend into more anterior motion-sensitive regions (40). This finding suggests that the OPA is situated near the highest stage of the occipito-parietal circuit before it diverges into multiple, parallel pathways in the parietal lobe (23). Thus, affordance representations in the OPA could underlie a number of navigational functions, including the planning of routes and online action guidance, but exactly what these functions are, and the downstream projections that support them, remain to be determined.

In conclusion, we have characterized a perceptual mechanism for identifying the navigational affordances of visual scenes. These affordances were encoded in the OPA, a scene-selective region of dorsal occipitoparietal cortex, and they appeared to be extracted automatically, even when subjects were not engaged in a navigational task. This visual mechanism appears to be well suited for providing information to the navigational system about the layout of paths and barriers in one's immediate surroundings.

Methods

Full details on materials and procedures are provided in SI Methods.

1. Gibson JJ (1958) Visually controlled locomotion and visual orientation in animals. *Br J Psychol* 49:182–194.
2. Gibson JJ (1977) The theory of affordances. *Perceiving, Acting and Knowing*, eds Shaw R, Bransford J (Erlbaum, Hillsdale, NJ).
3. Gallivan JP, Logan L, Wolpert DM, Flanagan JR (2016) Parallel specification of competing sensorimotor control policies for alternative action options. *Nat Neurosci* 19:320–326.
4. Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. *Annu Rev Neurosci* 33:269–298.
5. Greene MR, Oliva A (2009) Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognit Psychol* 58:137–176.
6. Greene MR, Baldassano C, Esteva A, Beck DM, Fei-Fei L (2016) Visual scenes are categorized by function. *J Exp Psychol Gen* 145:82–94.
7. Epstein RA, Higgins JS, Jablonski K, Feiler AM (2007) Visual scene processing in familiar and unfamiliar environments. *J Neurophysiol* 97:3670–3683.
8. Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601.
9. Maguire EA (2001) The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scand J Psychol* 42:225–238.
10. Grill-Spector K (2003) The neural basis of object perception. *Curr Opin Neurobiol* 13:159–166.
11. Epstein RA (2014) Neural systems for visual scene recognition. *Scene Vision: Making Sense of What We See*, eds Kveraga K, Bar M (MIT Press, Cambridge, MA), pp 105–134.
12. 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.
13. 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.
14. Walther DB, Chai B, Caddigan E, Beck DM, Fei-Fei L (2011) Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proc Natl Acad Sci USA* 108:9661–9666.
15. Cant JS, Xu Y (2012) Object ensemble processing in human anterior-medial ventral visual cortex. *J Neurosci* 32:7685–7700.
16. Harel A, Kravitz DJ, Baker CI (2013) Deconstructing visual scenes in cortex: Gradients of object and spatial layout information. *Cereb Cortex* 23:947–957.
17. Marchette SA, Vass LK, Ryan J, Epstein RA (2015) Outside looking in: Landmark generalization in the human navigational system. *J Neurosci* 35:14896–14908.
18. Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. *J Neurosci* 29:13992–14003.
19. Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. *J Neurophysiol* 106:2108–2119.
20. 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.
21. Tucker M, Ellis R (1998) On the relations between seen objects and components of potential actions. *J Exp Psychol Hum Percept Perform* 24:830–846.
22. Chao LL, Martin A (2000) Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12:478–484.
23. Kravitz DJ, Saleem KS, Baker CI, Mishkin M (2011) A new neural framework for visuospatial processing. *Nat Rev Neurosci* 12:217–230.
24. Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends Neurosci* 15:20–25.
25. Andersen RA, Buneo CA (2002) Intentional maps in posterior parietal cortex. *Annu Rev Neurosci* 25:189–220.
26. Dilks DD, Julian JB, Kubilius J, Spelke ES, Kanwisher N (2011) Mirror-image sensitivity and invariance in object and scene processing pathways. *J Neurosci* 31:11305–11312.
27. Dilks DD, Julian JB, Paunov AM, Kanwisher N (2013) The occipital place area is causally and selectively involved in scene perception. *J Neurosci* 33:1331–6a.
28. Ganaden RE, Mullin CR, Steeves JKE (2013) Transcranial magnetic stimulation to the transverse occipital sulcus affects scene but not object processing. *J Cogn Neurosci* 25:961–968.
29. Persichetti AS, Dilks DD (2016) Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex* 77:155–163.
30. Julian JB, Ryan J, Hamilton RH, Epstein RA (2016) The occipital place area is causally involved in representing environmental boundaries during navigation. *Curr Biol* 26:1104–1109.
31. Silson EH, Chan AW-Y, 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.
32. Epstein RA (2008) Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci* 12:388–396.
33. Auger SD, Mullally SL, Maguire EA (2012) Retrosplenial cortex codes for permanent landmarks. *PLoS One* 7:e43620.
34. 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.
35. Epstein RA (2005) The cortical basis of visual scene processing. *Vis Cogn* 12:954–978.
36. Epstein R, Graham KS, Downing PE (2003) Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* 37:865–876.
37. Epstein RA, Vass LK (2013) Neural systems for landmark-based wayfinding in humans. *Philos Trans R Soc Lond B Biol Sci* 369:20120533.
38. Janzen G, van Turennout M (2004) Selective neural representation of objects relevant for navigation. *Nat Neurosci* 7:673–677.
39. Viard A, Doeller CF, Hartley T, Bird CM, Burgess N (2011) Anterior hippocampus and goal-directed spatial decision making. *J Neurosci* 31:4613–4621.
40. Silson EH, Groen IIA, Kravitz DJ, Baker CI (2016) Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *J Vis* 16:14.
41. Harel J, Koch C, Perona P (2006) Graph-based visual saliency. *Advances in Neural Information Processing Systems 19*, eds Schölkopf B, Platt J, Hofmann T (MIT Press, Cambridge, MA), pp 545–552.
42. Aguirre GK (2007) Continuous carry-over designs for fMRI. *Neuroimage* 35:1480–1494.
43. 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.
44. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103:3863–3868.
45. Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum Brain Mapp* 15:1–25.

Supporting Information

Bonner and Epstein 10.1073/pnas.1618228114

SI Methods

Subjects. Twelve healthy subjects (8 female, mean age 24.3 y, SD = 2.9) from the University of Pennsylvania community participated in Exp. 1 and 16 (8 female, mean age 29.4 y, SD = 4.8) in Exp. 2. All participants 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. Two additional subjects were scanned in Exp. 1 and three in Exp. 2, but the data from these subjects were discarded before analysis because they reported difficulty staying awake during part of the experiment.

MRI Acquisition. Participants for Exp. 1 were scanned on a Siemens 3.0 T Trio scanner using a 32-channel head coil. We acquired T1-weighted structural images for anatomical localization using an MPRAGE protocol [repetition time (TR) = 1,620 ms, echo time (TE) = 3.09 ms, flip angle = 15°, matrix size = 192 × 256 × 160, voxel size = 1 × 1 × 1 mm]. We also acquired T2*-weighted functional images sensitive to BOLD contrasts using a gradient echo planar imaging (EPI) sequence (TR = 3,000 ms, TE = 30 ms, flip angle = 90°, matrix size = 64 × 64 × 44, voxel size = 3 × 3 × 3 mm).

Participants in Exp. 2 were scanned on a Siemens 3.0 T Prisma scanner using a 64-channel head coil. We acquired T1-weighted structural images using an MPRAGE protocol (TR = 2,200 ms, TE = 4.67 ms, flip angle = 8°, matrix size = 192 × 256 × 160, voxel size = 0.9 × 0.9 × 1 mm). We also acquired T2*-weighted functional images sensitive to BOLD contrasts using a multiband acquisition sequence (TR = 2,000 ms for main experimental scans and 3,000 ms for localizer scans, TE = 25 ms, flip angle = 70°, multiband factor = 3, matrix size = 96 × 96 × 81, voxel size = 2 × 2 × 2 mm).

Stimuli.

Exp. 1. We created images of artificially rendered indoor environments using the architectural design software Chief Architect. All environments were rectangular rooms of the same shape and size, with three walls visible, and 0–3 exits along the walls. There were eight experimental conditions in which navigational affordances were varied by altering the number and layout of these exits (Fig. 1). For each condition we created 18 aesthetic variants that differed in surface textures and the shapes of the doorways, and for each of these aesthetic variants, we created one stimulus in which walls with no exit were blank and one stimulus in which walls with no exit contained an abstract painting. (Note that the painting manipulation did not apply to the three-door condition in which all walls contained an exit.) The paintings were similar in size and shape to the exits and were located in the center of the wall where an exit would otherwise be. This resulted in 36 unique stimuli for each condition except for the three-door condition, which contained 18 unique stimuli (270 stimuli total). All images were 1,203 × 828 pixels and subtended a visual angle of ~18.5° × 12.8°.

Exp. 2. Stimuli were 50 color photographs depicting indoor environments viewed at eye level with clear navigational paths passing through the bottom of the image. Indoor environments were used because the navigational paths in interior spaces are typically more restricted and thus easier to define than the navigational paths in outdoor spaces; however, we have no theoretical reason to believe that our results would not generalize to outdoor environments. All images were 1,024 × 768 pixels and subtended a visual angle of ~17.1° × 12.9°.

The 50 images used in the experiment were selected from a larger set of 173 images. The principal navigational trajectories through each of these 173 images were established in a norming study in which 11 college-age participants used a computer mouse to indicate the paths that they would take to walk through each scene. Subjects began each path by clicking the mouse in a small red square at the bottom center of the image and then they moved the mouse along the path until clicking again to indicate the end of the path (Fig. 3A). After the second mouse click, the subjects were given the option of deleting and redrawing the previous path or moving on. The subjects were instructed to only draw visible paths on the ground floor without drawing on stairs or behind barriers where paths could potentially be inferred, and they were instructed to draw as many paths as they saw in each scene before moving on to the next stimulus. The responses were summed across subjects to create a heat map of the navigational paths in each scene (Fig. 3A).

We then modeled these navigational affordances using a set of hypothesized affordance-encoding channels that broadly capture the potential navigational paths to the left, center, and right of each scene (Fig. 3B). This procedure was inspired by previous work using a similar modeling approach to characterize the coding of color and orientation (18, 19). The underlying assumptions of the model are: (*i*) that the visual system contains clusters of affordance-selective neurons that can be characterized by neural tuning curves over a range of angular directions, and (*ii*) voxel responses have an approximately linear relationship to the summed responses of the neurons that they sample. Each of the affordance-encoding channels in our model was characterized by a tuning curve over a range of angles radiating from the bottom center of the image. These tuning curves were created by calculating half-wave rectified and squared sinusoids as a function of affordance angle. We generated a set of six equally spaced sinusoidal functions spanning the range of 0–360° and then restricted the model to the three functions with peaks in the range of 0–180° (i.e., the visible range in the images). To compute the responses of the affordance-encoding channels, we first needed to summarize the navigational-trajectory data over the range of angles from 0–180°. We began by mapping the pixel space of the stimulus images to polar coordinates, with the origin set at the bottom center of the image (Fig. 3A, *Right*). From this, we generated a set of 1° angular bins in which each pixel was labeled according to the bin that it fell within. We then summed the pixel values of the trajectory heat maps within each bin to produce angular histograms of the trajectory data (Fig. 3B). We were then able to generate the representations of our model by taking the dot products of these angular-histogram vectors and the tuning curves for the affordance-encoding channels.

We selected the 50 images used in the fMRI experiment with the goal that the navigational affordances in this set would be minimally correlated with low-level visual features and attentional saliency. We ran each of the images through two models that capture low-level visual properties of the scenes: edge detection and the Gist descriptor (20). We used edge detection to ensure that our model of navigational affordances was not confounded with a first-order representation of the lengths and orientations of contours in the images. Similarly, we used the Gist descriptor to ensure that our model of navigational affordances did not simply reflect the coarse spatial structure of the scenes. We also ran the images through a model of visual saliency that has previously been shown to generate highly accurate predictions of human eye fixations in natural images (41). This model was used

to minimize the possibility of confounding navigational affordances with attentional saliency. Edge detection was performed using the Sobel method as implemented in the edge function in MATLAB. We generated Gist descriptors by dividing each image into a 4 × 4 grid and calculating the average filter energy of eight oriented Gabor filters over four different spatial scales using software available at people.csail.mit.edu/torralba/code/spatialenvelope/. Visual saliency maps were calculated using the algorithm for graph-based visual saliency in the software available at www.vision.caltech.edu/~harel/share/gbvs.php. We compared the information content of the models using RSA. Specifically, we calculated the Spearman correlation of RDMs for the navigational-affordance model and each of the control models. The RDM for each model was created by vectorizing the model outputs for each image and then calculating the Pearson distance for all pairwise comparisons of model vectors. For the 50 images used in the fMRI experiment, the correlations of the navigational-affordance model with each of the computer-vision models were not reliably greater than chance (P values were obtained from an approximate permutation distribution as implemented in the MATLAB function corr): edge-detection model ($r = 0.02$, $P = 0.53$), Gist model ($r = 0.05$, $P = 0.11$), and the visual saliency model ($r = -0.02$, $P = 0.59$).

Procedure.

Exp. 1. Subjects were scanned with fMRI while viewing images of artificially rendered rooms and performing a color-detection task. On each experimental trial, an image appeared on the screen for 2 s followed by a 1-s interstimulus interval. Subjects were asked to fixate on a cross that remained on the screen throughout the scan and to indicate by button press whether two dots appearing on the walls of each scene were the same color or different colors (Fig. S1). Successful performance on this task required subjects to attend to each scene but did not require them to plan a route through the scene or explicitly encode its affordance structure. The positions of the dots were matched across the eight experimental conditions to control for spatial attention, and the colors were matched across conditions to control the number of "same" and "different" responses.

The stimuli were divided into four sets of 72 images each, which were presented in separate scan runs. These sets were created by first splitting the stimuli for each of the eight experimental conditions by aesthetic variant (renderings 1–9 vs. renderings 10–18), and then further splitting scenes that contained paintings from scenes that did not. Thus, each scan run contained nine exemplars for each of the eight experimental conditions. In addition to these stimuli, each scan run included nine null trials in which subjects viewed the fixation cross for 6 s and made no response. The final stimulus in the carryover sequence was also displayed in the first experimental trial of the scan so that the carryover effect for this stimulus "wrapped around" to the start of the scan. This first trial was not modeled in the statistical analyses. Each scan began with five 3-s null trials that were discarded and ended with an additional four 3-s null trials to capture the lag of the hemodynamic response to the final stimulus. This made for a total length of 5 min for each scan run. Stimuli within each scan run were presented in a type 1 index 1 continuous-carryover sequence so that each condition preceded and followed every other condition (including the null condition) an equal number of times (42). A unique continuous-carryover sequence was used for each run, and a unique set of sequences was generated for each subject. The four runs were presented twice for each subject, resulting in a total of eight experimental runs. The order of the runs was varied across subjects, subject to the constraint that all stimuli were shown once before presenting repeats and runs with paintings alternated with runs without paintings.

Exp. 2. Subjects were scanned with fMRI while viewing naturalistic scene images and performing a category-detection task. On each experimental trial, a scene appeared on the screen for 1.5 s followed

by a 2.5-s interstimulus interval. Subjects were asked to fixate on a cross that remained on the screen at all times and to press a button if the scene they were viewing was a bathroom. This task required subjects to attend to each scene and to process its meaning, but it did not require them to plan a route through the scene or explicitly encode its affordance structure. There were 10 scan runs of 4.8 min in length; in each, subjects viewed all 50 experimental stimuli (none of which were bathrooms), plus 5 bathroom images, 7 filler non-bathroom scenes (not from the experimental set), and 5 4-s null trials. The end of the scan included an additional 10 2-s null trials to capture the full hemodynamic response of the final trial. The bathroom and filler images were unique to each scan run. To minimize possible carryover effects associated with the task response, the trial order was constrained so that across runs all experimental images were preceded by a bathroom image an equal number of times; furthermore, to reduce possible attentional effects from the start of the run on the signals for the experimental trials, the first two trials of each run were always nonbathroom filler scenes. Beyond these constraints, the order of stimuli was randomized for each run.

fMRI Preprocessing and Modeling. fMRI data were processed and modeled using SPM12 (Wellcome Trust Centre for Neuroimaging) and MATLAB (R2015a Mathworks). All functional images for each participant were realigned to the first image and coregistered to the structural image. No smoothing was applied to these images. Voxelwise responses to each condition in each experiment were estimated using a general linear model. In Exp. 1, the model included regressors for each of the eight affordance conditions in each of the four stimulus sets (32 total) and a regressor for each run. In Exp. 2, the model included regressors for each experimental image (50 total), and regressors for the bathroom trials, the novel-scene trials, the start-up trials, and each run. Each regressor consisted of a series of impulse functions convolved with a standard hemodynamic response function. Parameter estimates for repeated conditions were averaged across runs. Low-frequency drifts were removed using a high-pass filter with a cutoff period of 128 s, and temporal autocorrelations were modeled with a first-order autoregressive model.

Functional Localizer and ROIs. BOLD responses in two additional scan runs were used to define the functional ROIs. In the first experiment, these two runs occurred at the end after the main experimental runs. In the second experiment, one localizer run occurred in the middle of the main experimental runs (to break up the repetitiveness of the experiment) and the other occurred at the end. We used a standard localizer protocol that included 16-s blocks of scenes, faces, objects, and scrambled objects. Images were presented for 600 ms with a 400-ms interstimulus interval, and subjects performed a one-back repetition-detection task on the images.

Images from the functional localizer scans were smoothed with an isotropic Gaussian kernel (3-mm FWHM). We identified the three scene-selective ROIs: the PPA, OPA, and RSC in each subject using a contrast of scenes > objects implemented in a general linear model and a group-based anatomical constraint of scene-selective activation derived from a large number ($n = 42$) of localizer subjects in our laboratory (17, 43). Using a similar approach, we defined an ROI for the EVC with the contrast of scrambled objects > scenes. Bilateral ROIs were identified using the top voxels within each hemisphere for the localizer contrasts (i.e., the difference in β-values across conditions). We created ROIs with a similar volumetric size in both experiments (~1,600 μL). This corresponded to 60 voxels per bilateral ROI in the first experiment and 200 voxels per bilateral ROI in the second experiment.

RSA. We used RSA to characterize the information encoded in the population responses of ROIs. First, we constructed neural

RDMs for each ROI through pairwise comparisons of the multivoxel activation patterns for each condition. Neural dissimilarities were calculated using Pearson distance (i.e., one minus the correlation coefficient). We then measured RSA fits for each subject as the Spearman correlation between these neural RDMs and model RDMs of representational content. Significance was assessed using one-tailed t tests of the RSA fits across subjects.
Exp. 1. In the first experiment, we constructed neural RDMs by performing pairwise comparisons of conditions across runs. The patterns being compared always corresponded to stimuli that differed on the aesthetic variants of the scenes and could also differ on the presence or absence of paintings (Fig. S2). We performed three kinds of cross-run comparisons. First, we compared patterns elicited by the two stimulus sets containing paintings. Second, we compared patterns elicited by the two stimulus sets that did not contain paintings. Third, we compared patterns elicited by the two stimulus sets containing paintings to patterns elicited by the two stimulus sets not containing paintings. The use of cross-run comparisons allowed us to test for representations of navigational affordances that were tolerant to the other visual properties of the stimulus sets. In these cross-run RDMs, the on-diagonal elements are meaningfully defined and are thus included in the RSA models. Off-diagonal elements for comparisons of the same two conditions were averaged together.

To construct a model RDM of navigational affordances, we first represented each condition using a 3D binary vector that codes for the presence of an exit along the left, center, and right walls. For example, an exit on the left was represented by the vector: [1 0 0]. We then calculated the Hamming distance (i.e., the percentage of coordinates that differ) for all pairwise comparisons of these condition vectors. This model RDM was then compared with all three categories of cross-run neural RDMs. The individual RSA effects for each category of cross-run comparisons are plotted in Fig. S2 and the average of these effects is plotted in Fig. 2.
Exp. 2. In the second experiment, we constructed a single neural RDM for each ROI through pairwise comparisons of the activation patterns for the 50 experimental images. We compared this neural RDM with a model RDM of navigational affordances based on responses from a norming study (this model is described in detail in the *Stimuli* section above).

Model Comparison. In Exp. 2, we compared the RSA effects for the navigational affordance RDM to the RDM for the Gist computer vision model. Using partial Spearman correlations, we identified the RSA correlations that could be uniquely attributed to each model RDM after partialling out the variance of the other RDM. We also performed an analysis to examine the correlation of these models in scene-selective ROIs after partialling out the lower-level representations of the EVC. To do this, we used the same partial correlation approach described above, but additionally partialled out the variance of the EVC RDM in each scene-selective ROI within each subject. This process allowed us to identify RSA effects that were present in scene-selective ROIs but not in the lower-level EVC.
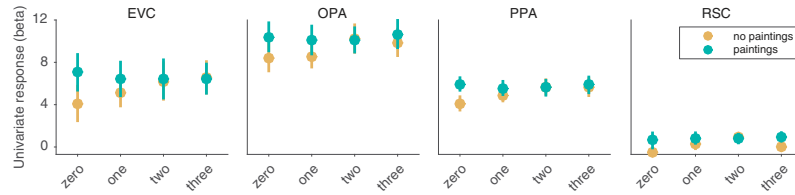
Searchlight Analyses. We performed whole-brain searchlight analyses to test for possible effects outside of our ROIs (44). For these analyses, we performed the same RSA procedures described above using multivoxel patterns within searchlight spheres (6-mm radius) centered at each voxel. RSA fits were assigned to the voxel at each searchlight's center, producing whole-brain maps of locally multivariate information coding. These images were smoothed with an isotropic Gaussian kernel (8-mm FWHM), normalized to standard Montreal Neurological Institute space, and submitted to whole-brain voxelwise t tests of random effects across subjects. To find the true type I error rate for each searchlight analysis, we performed Monte Carlo simulations, which involved 5,000 sign

permutations of the whole-brain data from individual subjects (45). We then reported voxels that were significant at $P < 0.05$ after correcting for whole-brain FWE.
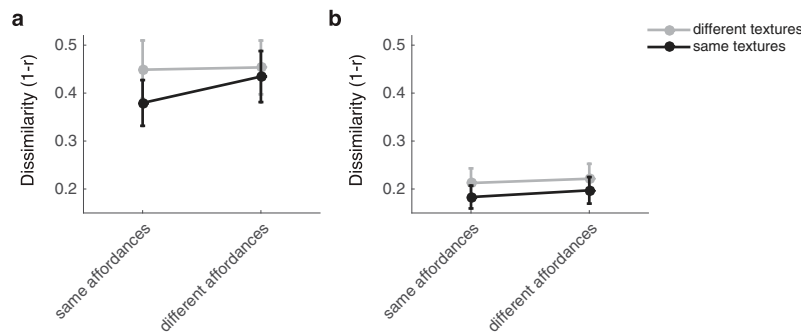
ANOVA of Affordances and Textures. In addition to the RSA approach, we also performed 2×2 repeated-measures ANOVAs on pattern similarity values in Exp. 1 to examine the influence of aesthetic features on the coding of affordances in the OPA and PPA. Specifically, for each subject we computed the median neural dissimilarity for four categories of comparisons based on two levels (same or different) of the following two factors: affordances and aesthetic variants (i.e., textural features and the shapes of the doorways). These neural dissimilarity measures were calculated from cross-run comparisons in which one stimulus set contained paintings and the other did not. The motivation for this approach was that the hypothesized coding of landmark (or place) identity in the PPA would be sensitive to stable environmental features, such as surface textures and structural properties, but would be relatively tolerant to less-stable features, such as paintings along the walls. The results of this analysis are summarized in Fig. S4.

Reconstruction Model. We performed a reconstruction analysis on the data from Exp. 2 to test the possibility of mapping the navigational affordances of novel scenes from the activation patterns in each ROI. We specifically examined the possibility of using ROI activation patterns to reconstruct a 2D heat map of the navigational paths in a scene. We first smoothed the response data from the norming study to reduce the noise inherent in the path-drawing procedure and to allow blending together of nearby trajectories along the same path (Fig. 4*A*). Smoothing was performed by convolving the images with a symmetric, 2D Gaussian filter spanning 60 pixels by 60 pixels, with σ set at half of the filter width. All analyses discussed below were performed on these smoothed navigational heat maps. The fMRI data were split into even and odd runs so that model training could be performed in one half of the data (i.e., the estimation set) and model assessment in the other half (i.e., the validation set). The procedures described below were iterated so that each half of the data served as an estimation set on one iteration and a validation set on the other. The prediction results were averaged across iterations. To maximize the reconstruction accuracy of the model, we first concatenated the ROI responses across subjects to create a multisubject response matrix for each split-half dataset. We then reduced the dimensionality of these matrices through PCA. The data were z-scored within each set and PCA was performed on the multisubject response matrix of the estimation set. We retained 45 PCs in the estimation set and used the loadings for these components to generate PC scores for the validation set. The reconstruction results were not contingent on the specific number of PCs retained; we observed similar results across the range from 30 to 45 PCs. We used ordinary least-squares regression to predict the pixelwise values of the smoothed navigational heat maps from the multisubject ROI PCs. We assessed the ability of this model to reconstruct the affordances of novel scenes through LOO cross-validation of the 50 experimental images. On each fold of this cross-validation procedure, the parameter estimates of the regression model were computed on the ROI PCs of the estimation set for the 49 training images. These learned parameter estimates were then applied to the ROI PCs of the validation set to generate a predicted affordance map for the held-out image. To remove the baseline similarity of the conditions, which largely reflects the shared starting point of each path, we mean-centered the navigational heat maps relative to the mean of the 49 training images within each fold. This process allowed us to evaluate how well each ROI predicted the remaining variance across conditions. When visualizing the reconstructed heat maps, the mean values were added back to the data. Reconstruction accuracy was calculated

by taking the average correlation of the predicted and actual maps of the navigational trajectories for all images. We computed the chance level for reconstruction accuracy at $P < 0.05$ corrected for FWE through a permutation test in which condition labels were randomized and the maximum reconstruction accuracy was calculated across all ROIs in 5,000 iterations.

SI Searchlight Results

The peak searchlight effect in Exp. 2 was located just outside of the OPA parcel. The group-constrained OPA parcel reflects the territory showing consistent scene-selective responses across a large number of participants; thus, it is somewhat conservative in its anatomical extent. We therefore compared the results of the searchlight clusters to scene-selective activation in the region of the OPA without constraining it to the parcel (Fig. S9). The scene-selective activation extended beyond the boundaries of the parcel in both experiments,

and partly differed across studies. Interestingly, the pattern of diverging scene-selective activation across Exps. 1 and 2 roughly corresponded to the pattern of divergence in their searchlight effects for the affordance model. Specifically, the scene-selective activation and the searchlight effects extended more inferiorly in Exp. 1 (particularly in the left hemisphere) and more superiorly in Exp. 2 (particularly in the right hemisphere). Indeed, the searchlight cluster from Exp. 2 showed strong selectivity for scenes relative to objects [$t(15) = 3.52$ $P = 0.0015$], which was larger than the selectivity index for the OPA parcel as a whole and did not differ in strength from the scene selectivity in any of the ROI parcels (all $P > 0.087$ for direct comparisons between the searchlight cluster and the OPA, PPA, and RSC parcels). These results and the anatomic distribution of the unconstrained localizer effects support the conclusion that the region showing affordance coding in Exp. 2 is part of the OPA rather than being a separate, adjoining region.

Fig. S1. Examples of rendered environments from Exp. 1 with overlaid colored dots. On each trial, subjects indicated whether the colors of the two dots were the same or different. Positions and colors of the dots were matched across the eight affordance conditions.

Fig. S2. Coding of navigational affordances in artificially rendered environments is tolerant to visual differences. (*A*) Model RDM of navigational affordances. (*B*) RSA for scenes without paintings. The OPA, EVC, and RSC exhibited significant effects, perhaps reflecting low-level visual differences between the affordance conditions. (*C*) RSA for scenes with paintings. In these comparisons, low-level visual difference between the affordance conditions are reduced; nevertheless, strong coding for affordances is observed in the OPA. (*D*) RSA for scenes with and without paintings. In these comparisons, scenes with similar affordances are highly dissimilar with respect to low-level visual features, which accounts for the negative correlations in the EVC. Despite this, the coding of navigational affordances remained significant in the OPA. Error bars represent $\pm$ 1 SEM; *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$.

Fig. S3. fMRI response as a function of the number of doorways in a scene. Each panel shows the mean univariate response within an ROI for stimuli grouped by the number of doorways they contain. Results are plotted separately for scenes with and without paintings (i.e., half of the stimuli contain a painting on any wall without a doorway). For scenes without paintings, the univariate responses in the EVC, OPA, and PPA increased in relation to the number of doorways [repeated-measures F-test for a positive linear effect of doorways; EVC: $F_{(1, 11)} = 8.51$, $P = 0.014$; OPA: $F_{(1, 11)} = 8.50$, $P = 0.014$; PPA: $F_{(1, 11)} = 14.96$, $P = 0.0026$; RSC: $F_{(1, 11)} = 2.06$, $P = 0.18$]. However, this effect was not observed in any ROI for stimuli with paintings (all $P > 0.62$), which are better matched on visual complexity across conditions with different numbers of doorways. This finding suggests that the linear trend in the univariate responses to the no-painting stimuli likely reflects an increase in visual complexity with the increasing number of doorways, rather than a specific effect of the number of pathways. Error bars represent $\pm$ 1 SEM.

Fig. S4. Coding of navigational affordances and textural features in artificially rendered environments. (A) Cross-run measures of neural dissimilarity in the PPA. A repeated-measures ANOVA showed significant main effects of textures [$F_{(1, 11)} = 7.61$, $P = 0.019$] and affordances [$F_{(1, 11)} = 14.74$, $P = 0.003$] and a trend toward an interaction of these factors [$F_{(1, 11)} = 3.99$, $P = 0.071$]. Post hoc t tests showed that the effect of affordance coding was only significant for environments with the same textures [$t_{(11)} = 4.01$, $P = 0.001$] but not for environments with different textures [$t_{(11)} = 0.30$, $P = 0.385$]. (B) Cross-run measures of neural dissimilarity in the OPA. A repeated-measures ANOVA showed significant main effects of textures [$F_{(1, 11)} = 8.86$, $P = 0.013$] and affordances [$F_{(1, 11)} = 10.24$, $P = 0.008$] and no evidence for an interaction of these factors [$F_{(1, 11)} = 0.50$, $P = 0.494$]. Post hoc t tests showed that the effect of affordance coding was significant both for environments with the same textures [$t_{(11)} = 2.24$, $P = 0.023$] and for environments with different textures [$t_{(11)} = 2.52$, $P = 0.014$]. Error bars represent $\pm$ 1 SEM.
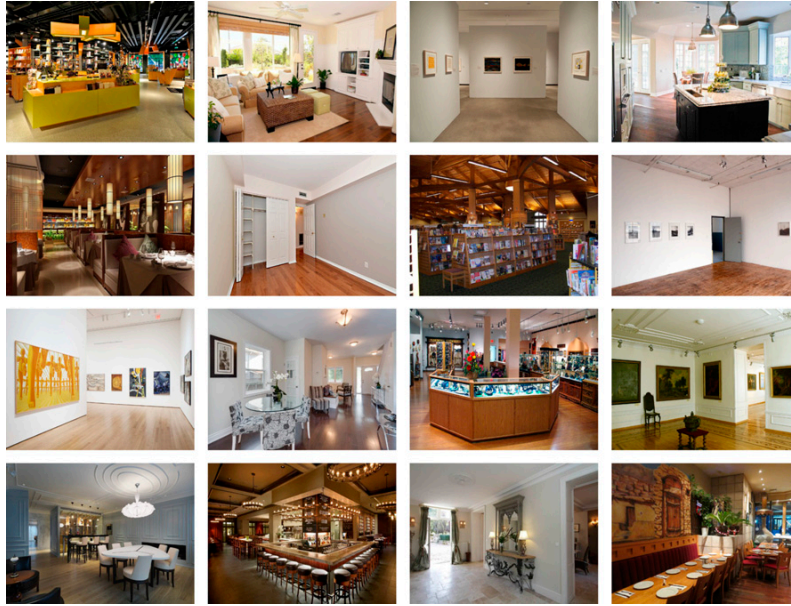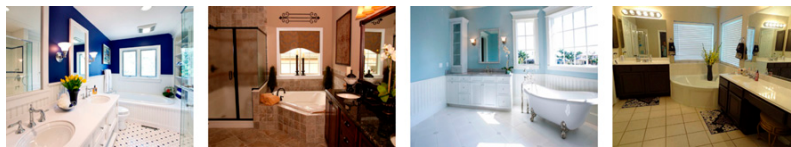
Fig. S5. Examples of complex, natural environments used as stimuli in Exp. 2. All images were eye-level photographs of indoor environments with clear navigational paths extending from the implied viewing position.


Fig. S6. Examples of target stimuli (i.e., bathroom scenes) in Exp. 2. Subjects were asked to monitor the semantic categories of the scenes and respond whenever the scene was a bathroom.

Fig. S7. Visualization of the navigational-affordance RDM. Using t-distributed stochastic neighbor embedding (t-SNE; perplexity set at 30), a 2D embedding was created that best captures that representational distances of the model RDM (shown in Fig. 3*C*). The *Left* panel shows how the stimulus images are arranged in this embedding. The plots on the *Right* illustrate the response level for each stimulus in the affordance channels of the encoding model (C, center; L, left; R, right; see also Fig. 3*B*). Colors are scaled to the minimum and maximum within each channel.
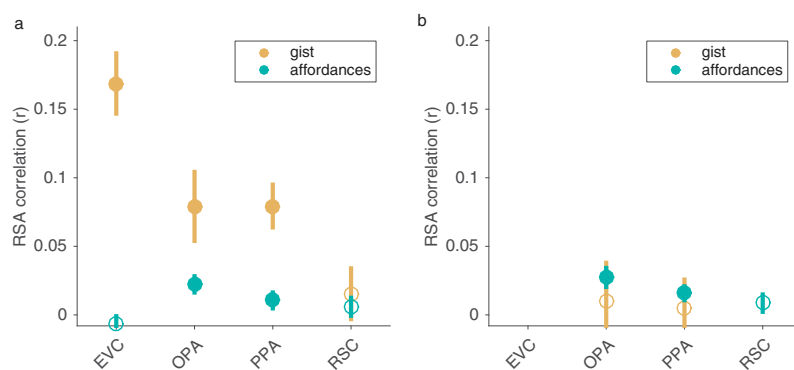
Fig. S8. Comparison of RSA effects for navigational affordances and Gist features. (*A*) Partial correlation analyses were used to identify the RSA correlations that could be uniquely attributed to each model. The RSA effects for the Gist model were strongest in EVC, whereas the effects of the affordance model were strongest in the OPA and PPA. (*B*) After partialling out the variance of the EVC RDM from each of the other ROIs, the effects of the affordance model remained significant in the OPA and PPA, but the effects of the Gist model dropped and were no longer significant. Error bars represent ± 1 SEM. Filled circles indicate significant effects at $P < 0.05$.
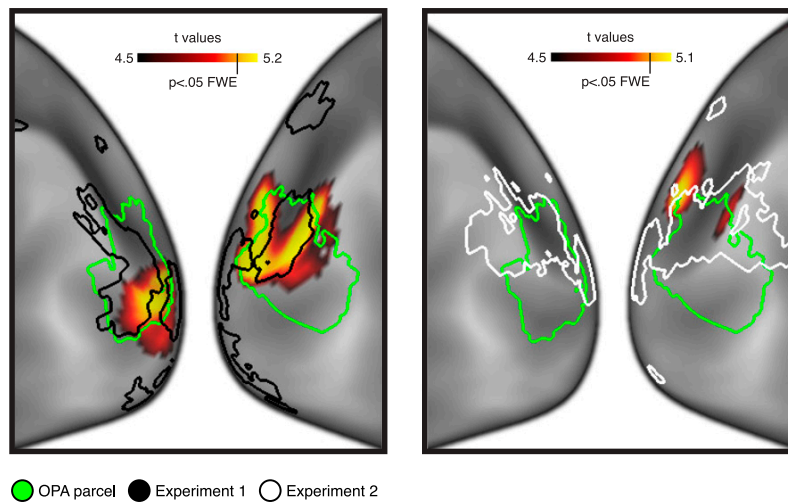
Fig. S9. Comparison of searchlight and localizer effects. These plots show close-up views of the searchlight results for the navigational-affordance model in Exp. 1 (*Left*) and Exp. 2 (*Right*) (see also Fig. 5). The overlaid outlines depict the OPA parcel (green) and localizer effects for the contrast of scenes > objects in Exps. 1 (black) and 2 (white). The localizer effects were thresholded at $P < 0.001$ uncorrected in a t test across subjects. These localizer clusters extend beyond the probabilistically defined OPA parcel, and they partly diverge across the two experiments. Furthermore, this pattern of divergence in the localizer effects across experiments roughly corresponds to the pattern of divergence in the searchlight effects. Specifically, the localizer and searchlight clusters extend more inferiorly in the subjects from Exp. 1 and more superiorly in the subjects from Exp. 2.