# Tracking the Emergence of Location-based Spatial Representations in Human Scene-Selective Cortex

Sam C. Berens[1,2] , Bárður H. Joensen[1,3,4], and Aidan J. Horner[1,5]

## Abstract

■ Scene-selective regions of the human brain form allocentric representations of locations in our environment. These representations are independent of heading direction and allow us to know where we are regardless of our direction of travel. However, we know little about how these location-based representations are formed. Using fMRI representational similarity analysis and linear mixed models, we tracked the emergence of location-based representations in scene-selective brain regions. We estimated patterns of activity for two distinct scenes, taken before and after participants learnt they were from the same location. During a learning phase, we presented participants with two types of panoramic videos: (1) an overlap video condition displaying two distinct scenes (0° and 180°) from the same location and (2) a no-overlap video displaying two distinct scenes from different locations (which served as a control condition). In the parahippocampal cortex (PHC) and retrosplenial cortex (RSC), representations of scenes from the same location became more similar to each other only after they had been shown in the overlap condition, suggesting the emergence of viewpoint-independent location-based representations. Whereas these representations emerged in the PHC regardless of task performance, RSC representations only emerged for locations where participants could behaviorally identify the two scenes as belonging to the same location. The results suggest that we can track the emergence of location-based representations in the PHC and RSC in a single fMRI experiment. Further, they support computational models that propose the RSC plays a key role in transforming viewpoint-independent representations into behaviorally relevant representations of specific viewpoints. ■

## INTRODUCTION

Rapidly learning the spatial layout of a new environment is a critical function that supports flexible navigation. This ability is thought to depend on the emergence of location-based representations in scene-selective brain regions that signal where we are irrespective of our current heading direction. As we are unable to sample all possible viewpoints from a given location simultaneously, the formation of location-based representations requires the integration of scenes from differing viewpoints. Despite evidence for the existence of location-based representations in scene-selective regions (e.g., Marchette, Vass, Ryan, & Epstein, 2015; Vass & Epstein, 2013), we know little about how such representations emerge.

Models of spatial navigation suggest that distinct brain regions are responsible for supporting allocentric (viewpoint-independent) and egocentric (viewpoint-dependent) representations of our environment (Julian, Keinath, Marchette, & Epstein, 2018; Byrne, Becker, & Burgess, 2007). Specifically, the parahippocampal cortex (PHC) and hippocampus are thought to encode allocentric spatial representations related to navigational landmarks/boundaries (Epstein, Patai, Julian, & Spiers, 2017; Burgess, Becker, King, & O'Keefe, 2001) and spatial context more broadly (Epstein & Vass, 2014). The hippocampus also supports a wider variety of spatial and nonspatial associative/configural functions in the service of memory and navigation (e.g., Henson & Gagnepain, 2010; Hannula & Ranganath, 2009; Kumaran et al., 2007; O'Keefe & Burgess, 2005; Eichenbaum, 2004). Here, we focus on the PHC given its more specific role in spatial allocentric processing relative to the hippocampus. In contrast, the parietal lobe is thought to support egocentric representations of specific viewpoints that underpin route planning (Calton & Taube, 2009; Byrne et al., 2007). To enable efficient route planning, a transformation between allocentric and egocentric representations is thought to occur in the retrosplenial cortex (RSC), cueing allocentric representations from egocentric inputs and vice versa (Bicanski & Burgess, 2018; Byrne et al., 2007).

In support of these models, human fMRI studies using representational similarity analyses (RSA) have found evidence for viewpoint-independent representations of specific locations (henceforth referred to as "location-based representations") in a network of brain regions including the PHC and RSC (Marchette, Vass, Ryan, & Epstein, 2014; Vass & Epstein, 2013). More recently, panoramic videos have been used to experimentally induce the formation of location-based representations (Robertson, Hermann, Mynick, Kravitz, & Kanwisher, 2016). Assessing pattern similarity for distinct scenes taken from the same location, Robertson et al. provided evidence for greater pattern similarity in the RSC and occipital place area (OPA) after

[1]University of York, [2]University of Sussex, [3]UCL Institute of Cognitive Neuroscience, [4]UCL Institute of Neurology, [5]York Biomedical Research Institute

participants had seen a panoramic video showing that two scenes were from the same location. This effect was not evident when participants could not learn that two scenes were from the same location. Interestingly, they also provided evidence for an effect in the PHC that occurred in both video conditions—that is, regardless of whether participants could learn the scenes were from the same location—suggesting a more general associative role for the PHC.

Despite these results, we still know little about (1) how quickly such representations are formed, (2) what types of spatial information they encode, and (3) under what conditions they are evoked. First, it remains unclear whether location-based representations emerge rapidly after short exposures to a new environment or whether they only develop after prolonged experience. Robertson et al. had participants watch videos outside the scanner, over the course of 2 days, before assessing pattern similarity inside the scanner. To test whether location-based representations can form rapidly, we developed a protocol that permitted us to scan participants before and after a short learning phase, allowing us to estimate changes in pattern similarity as a function of learning in a single fMRI experiment. Second, without tracking the formation of location-based representations, it is difficult to determine exactly what type of information they are representing. For instance, shared representations across viewpoints may relate to long-term semantic knowledge that is invoked when seeing different views of a well-known location (see Marchette, Ryan, & Epstein, 2017). In contrast, rapidly learning representations that are shared across different viewpoints of a new environment implies that the information being encoded is more likely to be spatial rather than semantic in nature.

Third, we do not know whether location-based representations are involuntary retrieved during visual processing. Computational models of spatial navigation predict that allocentric representations are automatically activated and updated by egocentric viewpoints (Bicanski & Burgess, 2018; Byrne et al., 2007). Furthermore, electrophysiological studies in rodents have shown that allocentric representations are automatically activated and updated during exploration (e.g., Monaco, Rao, Roth, & Knierim, 2014; O'Keefe & Dostrovsky, 1971). However, evidence in humans is lacking. Robertson et al. required participants to recall whether scenes were presented on the left or right of the screen, introducing a task that explicitly required them to recall the panorama, and the position of the specific scene within the panorama. Suggesting some level of involuntary retrieval, one fMRI study found that viewpoint-independent representations of specific buildings may be activated when participants judge whether the building is well known to them (Marchette et al., 2014). In the current study, participants performed an unrelated low-level attentional task as the scenes were presented. The activation of location-based representations under these conditions would suggest that they can be retrieved in a relatively automatic manner.

Here, we test whether location-based representations of novel environments can be learnt by integrating visual information across different scenes. Although location-based representations are predicted by models of spatial navigation, they may also be consistent with various other cognitive models (see Discussion). As such, we define location-based representations to be any type of information that encodes the relationship between different, nonoverlapping views of the same location. We recorded patterns of BOLD activity as participants passively observed a number of scenes depicting different views of novel locations. Subsequently, using an experimental manipulation introduced by Robertson et al. (2016), participants watched videos showing these scenes as part of a wider panorama. Half of the videos allowed participants to learn the spatial relationship between two scenes from the same location (overlap condition). The remaining videos acted as a control by presenting scenes from different locations (no-overlap condition). After the videos, we again recorded patterns of activity for each of the scenes. Whereas Robertson et al. (2016) only assessed scene representations after video presentation, we also scanned before and during the videos; see Clarke, Pell, Ranganath, and Tyler (2016) for a similar preexperimental versus postexperimental design focused on changes in object representations. This allowed us to track the potential emergence of location-based representations using RSAs as well as assess neural activity when these representations were being formed.

Using generalized linear mixed models, we show that patterns evoked by different scenes become more similar in scene-selective regions of the PHC and RSC after the presentation of the video panoramas. This increase in similarity was specific to the "overlap" video condition, where scenes from the same location were presented together, and was not observed in the no-overlap condition. This suggests the emergence of location-based representations in the PHC and RSC. Importantly, whereas this increase in pattern similarity emerged in the PHC regardless of behavioral performance, the same pattern was only present in the RSC when participants could remember which scenes came from the same location. This finding supports computational models that propose the RSC is critical in translating viewpoint-independent representations in the medial temporal lobe into more behaviorally relevant egocentric representations.

## METHODS

### Participants

Twenty-eight right-handed participants were recruited from the University of York, United Kingdom. These participants had no prior familiarity with the locations used as stimuli in the experiment (see below). All participants gave written informed consent and were reimbursed for their time. Participants had either normal or corrected-to-normal vision and reported no history of neurological or

psychiatric illness. Data from five participants could not be included in the final sample because of problems with fMRI data acquisition (one participant), excess of motion-related artifacts in the imaging data (three participants), and a failure to respond during one of the in-scanner tasks (one participant). As such, analyses included 23 participants (10 men) with a mean age of 21.96 years ($SD$ = 3.22 years). The study was approved by a local research ethics committee at the University of York.

## Stimuli

We generated 12 panoramic images of different urban locations from the City of Sunderland, and Middlesbrough town center, United Kingdom (Figure 1; osf.io/cgy97). These panoramas spanned a 210° field-of-view horizontally but were restricted in the vertical direction to limit the appearance of proximal features (<2 m from the camera). Throughout the experiment, 24 "endpoint images" displaying 30° scenes taken from either end of each panorama were shown (i.e., centered at 0° and 180°; Figure 1A). These images were shown both inside and outside the scanner to assess participants' spatial knowledge of the depicted locations and for the RSA (see below).

Endpoints were also shown in a series of videos (see osf.io/cgy97). In overlap videos, Images A1 and A2 (taken from opposite ends of the same panorama) were presented such that their spatial relationship could be inferred (Figure 1B). Here, a camera panned from each endpoint to the center of the panorama showing that A1 and A2

belonged to the same location. In contrast, a no-overlap video featured endpoints from two unrelated panoramas (Images B1 and C2). Again, these videos showed an end-to-center camera pan from each image. However, because there was no visual overlap between the video segments, observers could only infer that Endpoints B1 and C2 belonged to different locations. The no-overlap condition acted as a control condition, ensuring Endpoints B1 and C2 were seen in a similar video to endpoints in the overlap condition (A1 and A2), with the same overall exposure and temporal proximity. To ensure that the occurrence of a visual overlap was easily detectable, all videos alternated the end-to-center sweep from each endpoint over two repetitions.

Pairs of endpoints from the same panorama were grouped into sets of three. The first pair in each set was assigned to the overlap video condition (A1 and A2). Two endpoints from different panoramas were assigned to the no-overlap video condition (B1 and C2). The remaining endpoints belonged to an "unseen video" condition as they were not shown during any video (B2 and C1). These assignments were counterbalanced across participants such that each image appeared in all three conditions an equal number of times. The order of camera pans during videos (e.g., A1 first vs. A2 first) was also counterbalanced both within and across participants. Analyses showed the visual similarity of image endpoints was matched across experimental conditions as measured by the Gist descriptor (Oliva & Torralba, 2001) and local correlations in luminance and color information (osf.io/6sr9p/). Pilot data

**Figure 1.** Stimuli used during, and analyses performed on, the in-scanner tasks. (A) An example location panorama with two endpoint images. Single endpoints were shown during the in-scanner target detection task. As in Robertson et al. (2016), full panoramas were never shown as whole images but were presented during the in-scanner videos. (B) Depiction of the two video conditions: overlap versus no-overlap videos. Overlap videos showed camera pans from each endpoint of a given panorama (denoted A1 and A2) to the center of that panorama. The central overlap allowed participants to learn a spatially coherent representation that included both A1 and A2. No-overlap videos involved pans between endpoints from



two different locations (denoted B1 and C2), meaning that there was no visual overlap. (C) Similarity contrast matrices used to model changes in representational similarity between endpoints (i.e., between A1, A2, B1, B2, C1, and C2). Red squares indicate positively weighted correlations, and blue squares indicate zero-weighted correlations. From left to right, the matrices account for the representational similarity of endpoints: (1) from the same location regardless of video condition, (2) that were seen in the same video (including overlap and no-overlap videos), (3) in the unseen condition specifically, and (4) in the overlap condition specifically. Linear combinations of these matrices, along with their interactions with a session regressor (pre-videos vs. post-videos), accounted for each RSA effect across all experimental conditions.
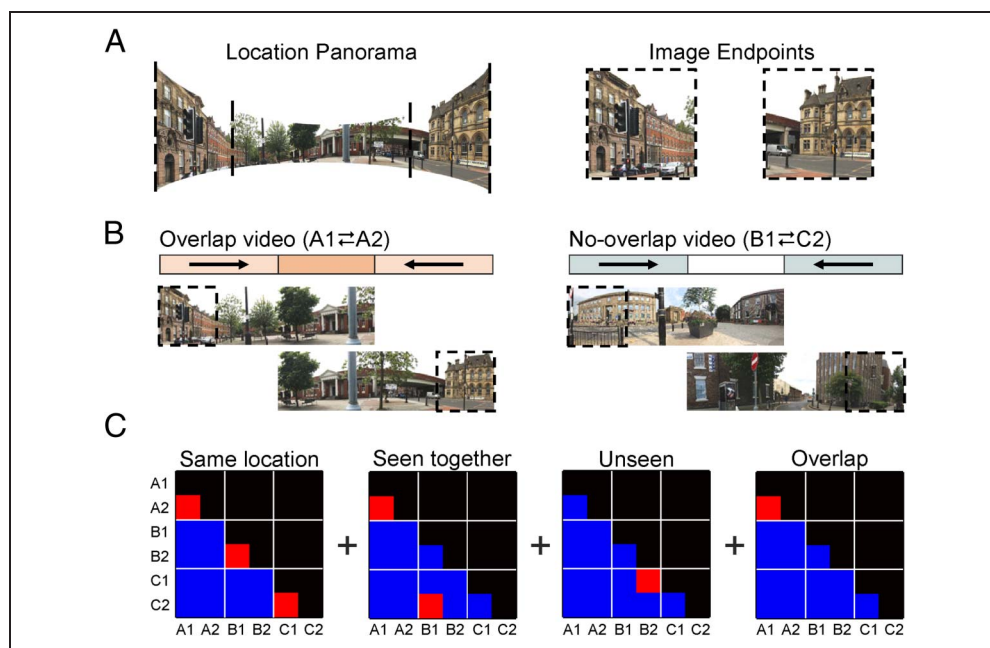
**Figure 2.** Behavioral task and results (A) A schematic illustration of the pre-video and post-video behavioral task. One endpoint was first presented as a cue (enclosed by red box), followed by five numbered alternatives. Participants were then prompted to select which one of the alternatives belonged to the same location as the cue. (B) Performance on the pre-video and post-video behavioral tasks plotted by video condition. Error bars represent 95% confidence intervals, and the dashed line at $p = .2$ reflects 5-AFC chance level.
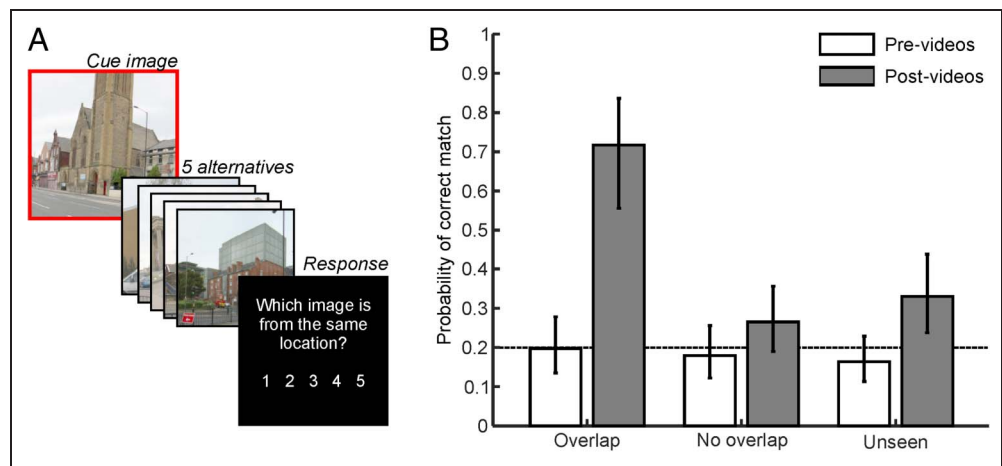
revealed that participants could not reliably identify which endpoints belonged to the same location without having seen the videos (osf.io/kgv64/).

## Procedure

Before entering the scanner, participants performed a behavioral task to assess their ability to infer which image endpoints were from the same location. Once in the scanner, they undertook a functional localizer task to identify scene-selective regions of the PHC and RSC. They were then shown each image endpoint multiple times (performing a low-level attentional task) to assess baseline representational similarity between each image endpoint (i.e., before learning). During a learning phase, overlap and no-overlap videos were presented, with participants instructed to identify whether the endpoints in each video belonged to the same location or not. After this video learning phase, each image endpoint was again presented multiple times to assess postlearning representational similarity between image endpoints. Finally, outside the scanner, participants performed the same behavioral task (as before scanning) to assess the extent to which participants had learnt which image endpoints belonged to the same location (and a further test of associative memory; see below). A figure illustrating the order and approximate duration of each experimental task is available at osf.io/zh8f2/.

### Prescanner/Postscanner Tasks

Participants were tested on their ability to identify which endpoints belonged to the same location both before and after scanning (both outside the scanner). On each trial, one endpoint surrounded by a red box was presented for 3 sec. After this, five other endpoints were displayed in a random sequence, each shown alongside a number denoting the order of appearance (i.e., 1–5; Figure 2A; 2 sec per image, 500-msec ISI). One image in the sequence (the target) was taken from the same panorama as the cue. The remaining four endpoints (lures) belonged to panoramas in the same set of stimuli. As such, if Endpoint B1 was presented as the cue, B2 would be the target, and A1, A2, C1, and C2 would be lures (i.e., a five-alternative forced choice, 5-AFC). After the five alternatives had been shown, participants were prompted to select the target using a numeric key press (1–5). Across 24 trials, each endpoint was used as a cue image.

After scanning and the second block of the location identity task described above, participants were also asked to identify which images appeared together in the same video. Note that this is slightly different to the previous task because participants could have known that Endpoints B1 and C2 appeared in the same video, despite not knowing which endpoints were from the same location (i.e., B2 and C1, respectively). Using a similar procedure to that described above, endpoints from either the overlap or no-overlap video conditions were cued and participants were asked to select the appropriate endpoint from the five alternatives in the same set.

### In-scanner Tasks

*Functional localizer.* Before the main experimental task, participants undertook a functional localizer scan with the purpose of identifying four scene-selective ROIs—in particular, the left and right PHC as well as the left and right RSC. This involved presenting four blocks of scene images (coasts, mountains, streets, and woodlands) interleaved with four blocks of face images (male and female). In each block, 10 unique images were shown in quick succession with a display time of 700 msec per image and an ISI of 200 msec. Blocks were separated with a 9-sec interblock interval, and their running order was counterbalanced across participants. The scene images used here were different to those in the main experiment, and none was repeated during the localizer itself. All images were shown in grayscale and were presented with a visual angle of ~14°. To ensure localizer images were being attended to, participants were tasked with detecting an oddball target that

was superimposed onto one of the images in each block. The target was a small red dot with a 3-pixel radius. When this was seen, participants were required to respond with a simple button press as quickly as possible (mean detection performance: $d' = 3.116$, $SD = 0.907$).

*Presentation of endpoint images.* Participants were shown all 24 endpoint images during an event-related functional imaging task. The task was optimized to measure multivariate patterns of BOLD activity specific to individual endpoints and was run both before and after participants had seen the panoramic videos (Session 1: pre-videos; Session 2: post-videos). All endpoints were presented nine times for both the pre-video and post-video functional run. Images were displayed for 2.5 sec with an ISI of 2 sec. The order of stimuli in each functional run was optimized to facilitate the decoding of unique BOLD patterns across endpoints (optimization algorithm available at osf.io/eh78w/). No image was presented on successive trials to avoid large adaptation effects, and the design included 12 null events in each functional run (i.e., 10% of all events). Like the functional localizer, participants were tasked with detecting an oddball target that was superimposed onto a small proportion of the images. Here, the target was a group of three small red dots (3-pixel radius, <0.2°), with each dot drawn at a random position on the image. Targets were present on one of every nine trials such that eight repetitions of each endpoint image were target free (target trials were not used to estimate BOLD patterns). As above, participants were required to respond to these targets with a simple button press (mean detection performance, $d'$, were 3.362 [$SD = 0.642$, pre-videos] and 3.659 [$SD = 0.485$, post-videos]).

*Panoramic video task.* Participants watched all video clips from the overlap and no-overlap video conditions while being scanned. Each video lasted 20 sec and was followed by a 10-sec rest period. In the first 3 sec of this rest period, participants were prompted to indicate whether each video segment depicted scenes from the same or different locations. Responses were recorded with a left/right button press. This question was asked to ensure that participants were attending to the visual overlap across segments (mean discrimination performance: $d' = 3.220$, $SD = 0.373$). All videos were repeated three times in a pseudorandom order to allow for sufficient learning. Before entering the scanner, participants were asked to remember which endpoints were seen together in the same video, even if they appeared in a no-overlap video. Participants were told that a test after the scan would assess their knowledge of this.

## MRI Acquisition

All functional and structural volumes were acquired on a 3-T Siemens MAGNETOM Prisma scanner equipped with a 64-channel phased array head coil. T2*-weighted scans were acquired with EPI, 35 axial slices (approximately 0° to the AC–PC line; interleaved), and the following parameters: repetition time = 2000 msec, echo time = 30 msec, flip angle = 80°, slice thickness = 3 mm, and in-plane resolution = 3 × 3 mm. The number of volumes acquired during (a) the functional localizer, (b) the video task, and (c) each run of the endpoint presentation task was 75, 363, and 274, respectively. To allow for T1 equilibrium, the first three EPI volumes were acquired before the task started and then discarded. Subsequently, a field map was captured to allow the correction of geometric distortions caused by field inhomogeneity (see the MRI Preprocessing section below). Finally, for purposes of coregistration and image normalization, a whole-brain T1-weighted structural scan was acquired with a 1-mm³ resolution using a magnetization prepared rapid gradient echo pulse sequence.

## MRI Preprocessing

Image preprocessing was performed in SPM12 (www.fil.ion.ucl.ac.uk/spm). This involved spatially realigning all EPI volumes to the first image in the time series. At the same time, images were corrected for field inhomogeneity based geometric distortions (as well as the interaction between motion and such distortions) using the Realign and Unwarp algorithms in SPM (Hutton et al., 2002; Andersson, Hutton, Ashburner, Turner, & Friston, 2001). For the RSA, multivariate BOLD patterns of interest were taken as $t$ statistics from a first-level general linear model (GLM) of unsmoothed EPI data in native space. Aside from regressors of interest, each first-level GLM included a set of nuisance regressors: six affine motion parameters, their first-order derivatives, regressors censoring periods of excessive motion (rotations > 1° and translations > 1 mm), and a Fourier basis set implementing a 1/128-Hz high-pass filter. For the analyses of univariate BOLD activations, EPI data were warped to Montreal Neurological Institute space with transformation parameters derived from structural scans (using the DARTEL toolbox; Ashburner, 2007). Subsequently, the EPI data were spatially smoothed with an isotropic 8-mm FWHM Gaussian kernel before GLM analysis (regressors included the same nuisance effects noted above).

## ROIs

We generated four binary masks per participant to represent each ROI in native space. To do this, a first-level GLM of the functional localizer data modeled BOLD responses to scene and face stimuli presented during the localizer task. Each ROI was then defined as the conjunction between a "scene > face" contrast and an anatomical mask of each region that had been warped to native space (left/right PHC sourced from Tzourio-Mazoyer et al., 2002; left/right RSC sourced from Julian, Fedorenko, Webster, & Kanwisher, 2012). Thus, the ROIs were functionally defined but

constrained to anatomical regions known to be spatially selective. Normalized group averages of these ROIs are available at osf.io/gbznp/ and neurovault.org/collections/4819.

Recent evidence suggests that the RSC is composed of at least two functionally distinct subregions, both of which may be scene selective: (1) a retinotopically organized medial place area in posterior sections of the RSC and (2) a more anterior region corresponding to BAs 29 and 30 associated with more integrative mnemonic processes (Silson, Steel, & Baker, 2016). In the current study, we focus on the functionally defined RSC as a whole and do not differentiate between these subregions. However, the functional ROIs that we identified for each participant principally cover anterior sections of the RSC corresponding to BAs 29 and 30 and show little overlap with the retinotopic areas identified by Silson et al.

The OPA has also been implicated as a critical scene-selective region (e.g., Robertson et al., 2016; Marchette et al., 2015). Recent research suggests that this region is principally involved in representing environmental boundaries and navigable paths during visual perception (Malcolm, Silson, Henry, & Baker, 2018; Bonner & Epstein, 2017; Julian, Ryan, Hamilton, & Epstein, 2016). However, computational models of spatial navigation do not predict that the OPA maintains location-based representations that are viewpoint invariant (Bicanski & Burgess, 2018; Byrne et al., 2007). In addition, we were only able to reliably delineate the OPA bilaterally in 6 of the 23 participants in our sample. As such, we did not focus on this region in the current study; instead, we restricted our main analyses and family-wise error (FWE) corrections to the PHC and RSC bilaterally. Nonetheless, for completeness, we generated an OPA mask using a normalized group-level contrast and ran the location-based RSA analyses reported below on this region separately (statistical outputs available at osf.io/d8ucj/). No effects of interest were identified in either the left or right OPA.

## RSAs

Our general approach to the RSA involved modeling the observed similarity between different BOLD patterns as a linear combination of effects of interest and nuisance variables. Here, the similarity between BOLD responses was taken as the correlation of normalized voxel intensities ($t$ statistics) across all voxels in an ROI. The resulting correlation coefficients were then Fisher-transformed before being subjected to statistical analysis. This transform ensures that the sampling distribution of similarity scores is approximately normal to meet the assumption of normality for statistical inference. We then entered all the transformed similarity scores under test from each participant and stimulus set into a general linear mixed-effects regression model. Although underused in the neurosciences (although see Motley et al., 2018), these models are common in the psychological literature as

they offer a robust method of modeling nonindependent observations with few statistical assumptions (Baayen, Davidson, & Bates, 2008). Here, we used mixed-effects models to predict observed representational similarity between endpoints with a set of fixed-effects and random-effects predictors (discussed below).

Importantly, mixed-effects models allow us to include estimates of pattern similarity across individual items (endpoints) and participants in the same statistical model. The fixed-effects predictors in each model specified key hypotheses of interest. The random effects accounted for statistical dependencies between related observations at both the item and participant levels. RSAs of fMRI data typically either assess patterns across all items (regardless of condition) or average across items in the same condition, meaning that important variation within conditions is ignored. Our modeling approach allows us to examine changes in representational similarity at the level of both items and conditions simultaneously while controlling for statistical dependencies between related observations.

Raw similarity data and mean similarity matrices are available on the Open Science Framework (osf.io/cgy97). This page also includes MATLAB functions for estimating each statistical model as well as the model outputs.

### Visual Representations of Specific Endpoints

We first examined whether the passive viewing of endpoint images evoked stimulus-specific visual representations in each of our four ROIs (left and right PHC and RSC). Multivariate BOLD responses to the endpoints were estimated for Session 1 (pre-videos) and Session 2 (post-videos) separately. We then computed the similarity of these responses across sessions by correlating BOLD patterns in Session 1 with patterns in Session 2. This resulted in a nonsymmetric, $24 \times 24$ correlation matrix representing the similarity between all BOLD patterns observed in Session 1 and those observed in Session 2. The correlation coefficients ($n = 576$ per participant) were then Fisher-transformed and entered as a dependent variable into a mixed-effects regression model with random effects for participants and endpoints. The main predictor of interest was a fixed effect that contrasted correlations between the same endpoints (e.g., A1–A1, B1–B1; $n = 24$ per participant) with correlations between different endpoints (e.g., A1–A2, A1–B1; $n = 552$ per participant) across the two sessions.

As well as running this analysis in each ROI, we performed a complementary searchlight analysis to detect endpoint-specific representations in other brain regions. Here, local pattern similarity was computed for each brain voxel using spherical searchlights with a 3-voxel radius (the mean number of voxels per searchlight was 105.56; searchlights were not masked by gray-/white-matter tissue probability maps). Fisher-transformed correlations for same versus different endpoints were contrasted at the first level before running a group-level random-effects analysis.

*Location-based Memory Representations*

We next tested our principal hypothesis—whether representations of Endpoints A1 and A2 became more similar to one another as a result of watching the overlap videos—in each ROI. Using the multivariate BOLD responses from Sessions 1 and 2, we computed the neural similarity between endpoints that were presented within the same image set and the same session. This resulted in eight symmetric, $6 \times 6$ correlation matrices for each participant—one per set in Session 1 and one per set in Session 2. All the correlation coefficients from the lower triangle of these matrices ($n = 15$) were then Fisher-transformed and entered as a dependent variable into a mixed-effects regression model (see Figure 1C). As such, the model included 120 correlation coefficients per participant (2 sessions $\times$ 4 sets $\times$ 15 similarity scores).

One fixed-effects predictor modeled unspecific changes in similarity between sessions (hereafter referred to the session effect) by coding whether similarity scores were recorded in Session 1 or Session 2. Similarly, a further three fixed-effects predictors modeled similarity differences attributable to (1) endpoints in the overlap condition (i.e., A1–A2), (2) endpoints shown in the same video (A1–A2, B1–C2), and (3) endpoints that were not shown in any video (C1–B2)—shown in Figure 1C. Together, these predictors and their interactions constituted a $2 \times 3$ factorial structure (Session [1 vs. 2] $\times$ Condition [overlap vs. no-overlap vs. unseen]) and so were tested with a Session $\times$ Condition $F$ test. Nonetheless, our principal hypothesis holds that there will be a specific interaction between the Session and Overlap predictors (referred to as the Session $\times$ Overlap effect), which we report alongside the $F$ test. The model also included a predictor indicating whether endpoints were from the same location (A1–A2, B1–B2, C1–C2), thereby allowing us to estimate changes in similarity between them. This ensured that variance loading onto the Session $\times$ Overlap effect was properly attributable to the learning of spatially coherent representations rather than some combination of other factors (e.g., same location + seen in the same video). Note that this model term quantifies similarity differences between overlap endpoints and all other endpoints that "change" between Session 1 and Session 2. A positive effect may indicate either an increase in similarity in the overlap condition or a decrease across all other similarity scores regardless of condition (or both). As such, the model is structured to account for any systematic change in the baseline level of similarity across sessions (see Results). Furthermore, the Session $\times$ Overlap term is only sensitive to a learning effect that causes relative shifts in similarity scores specific to the overlap condition and cannot be attributed to any other combination of effects.

Finally, the model included a behavioral predictor specifying whether participants were able to match Endpoints A1–A2 in the postscanner task (mean centered with three levels: 0, 1, or 2 correct responses per pair). This examined

whether changes in representational similarity were dependent on participants' ability to identify that endpoints from the overlap condition belonged to the same location after scanning (i.e., a three-way interaction: Session $\times$ Overlap $\times$ Behavior). Random effects in the model accounted for statistical dependencies across image sets, sessions, and participants.

To complement the ROI analyses, we ran a searchlight analysis that tested for RSA effects across the whole brain (searchlight radius: 3 voxels). Here, first-level contrast estimates compared the Fisher-transformed correlations between overlap endpoints (i.e., A1–A2) and all other endpoint correlations (e.g., B1–B2, B1–C1). A group-level analysis then compared these similarity contrasts between sessions to test the Session $\times$ Overlap interaction. To test for a Session $\times$ Overlap $\times$ Behavior interaction, the group-level model also included a behavioral predictor specifying a participant's average performance in matching A1 to A2 during the postscanner task (mean centered). Note that this searchlight analysis is not able to control for the potential contributions of other important factors (i.e., same location, same video) that our mixed-effects approach explicitly controls. It is complementary, but secondary, to the ROI analyses.

## Statistical Validation and Inference

To ensure that each mixed-effects regression model was not unduly influenced by outlying data points, we systematically excluded observations that produced unexpectedly large residuals more than 2.5 *SD*s above or below model estimates. This was conducted regardless of condition and so did not bias the analyses to finding an effect (if no effect were present). Furthermore, a highly similar pattern of results was seen when not excluding outliers, supporting the robustness of our findings (see osf.io/dzy3p). Following these exclusions, Kolmogorov–Smirnov tests indicated that residuals were normally distributed across all the linear mixed-effects models. In addition, visual inspection of scatterplots showing residual versus predicted scores indicated no evidence of heteroscedasticity or nonlinearity. Where effects size estimates are contrasted across different models, we report the result as an unequal variance *t* test with the degrees of freedom being approximated using the Welch–Satterthwaite equation (Welch, 1947).

All *p* values are reported as two-tailed statistics. FWE corrections related to the multiple comparisons across our four ROIs are made for each a priori hypothesis (denoted $p_{\mathrm{FWE}}$). In addition, we report whole-brain effects from searchlight and mass univariate analyses when they survive FWE-corrected thresholds ($p_{\mathrm{FWE}} < .05$) at the cluster level (cluster-defining threshold: $p < .001$ uncorrected). All other *p* values are noted at uncorrected levels. As well as reporting null hypothesis significance tests, we present the results of complimentary Bayesian analyses. Unlike the frequentist statistics, these indicate whether the null is statistically preferred over the alternative hypothesis.

As such, we use the Bayesian analyses to determine whether there is evidence for the null when frequentist tests are nonsignificant. For each $t$ test, a Bayes factor in favor of the null hypothesis ($BF_{01}$) was computed with a Cauchy prior centered at zero (i.e., no effect) and a scale parameter ($r$) of $\sqrt{0.5}$ (see Gelman, Jakulin, Pittau, & Su, 2008). Bayes factors greater than 3 are taken as evidence in favor of the null hypothesis, whereas those less than 1/3 are taken as evidence in favor of the alternative (Kass & Raftery, 1995). Finally, alongside the inferential statistics, we report Cohen's $d$ effect sizes for each $t$ test. When effects are tested in the context of a mixed-effects model, estimates of Cohen's $d$ are computed from the fixed effects only and exclude variance attributed to random effects.

## RESULTS

### Behavioral Performance

We first analyzed behavioral responses to the prescanner and postscanner tasks to determine (a) whether participants were able to identify which endpoints belonged to the same location and (b) whether performance increased as a result of watching the overlap videos. A generalized linear mixed-effects analysis modeled correct versus incorrect matches between cue and target endpoints as a function of Session (pre-videos vs. post-videos) and Experimental Condition (overlap, no overlap, and unseen). As such, the model constituted a $2 \times 3$ factorial design with random intercepts and slopes for both participants and endpoints.

The results, displayed in Figure 2B, revealed significant main effects of Session, $F(1, 1098) = 47.302, p < .001$, and Condition, $F(2, 1098) = 6.500, p = .002$, as well as an interaction between them, $F(2, 1098) = 11.231, p < .001$. The interaction indicated that performance was at chance level across all conditions before the videos (min: $p = .249, BF_{01} = 2.531, d = -0.241$) but substantially increased in the overlap video condition after the videos, $t(1098) = 6.867, p < .001, BF_{01} < 0.001, d = 1.432$ (post-video > pre-video). This increase was not seen in the no-overlap condition, $t(1098) = 1.761, p = .079, BF_{01} = 1.212, d = 0.3672$; however, a significant increase was seen in the unseen condition, $t(1098) = 3.159, p = .002, BF_{01} = 0.105, d = 0.659$. The performance increases in the control conditions (only significant in the unseen condition) were likely the result of participants being able to exclude overlap endpoints as nontarget alternatives in the 5-AFC test (i.e., a recall-to-reject strategy, disregarding A1 and A2 when cued with either B1, B2, C1, or C2). Consistent with this, Session 2 performance in the no-overlap and unseen conditions was not significantly different from chance level in a 3-AFC test (0.33, as opposed to 0.2 in a 5-AFC test; no-overlap: $t(1098) = -1.494, p = .135, BF_{01} = 1.729, d = -0.312$; unseen: $t(1098) = -0.054, p = .957, BF_{01} = 4.567, d = -0.011$). Nonetheless, performance in the overlap condition did significantly differ from this

adjusted chance level, $t(1098) = 4.514, p < .001, BF_{01} = 0.006, d = 0.941$.

Participants' increased ability to match endpoints in the overlap condition was not characteristic of a general tendency to match endpoints that appeared in the same video (i.e., selecting B1 when cued with C2). This was evident because matches between no-overlap endpoints were not more likely in Session 2 compared with Session 1, $t(366) = 0.646, p = .519, BF_{01} = 3.785, d = 0.135$. In contrast, performance increases in the overlap condition (i.e., the post-video > pre-video effect reported above) were significantly larger than this general effect of matching all endpoints that appeared in the same video, $t(949.20) = 5.027, p < .001, BF_{01} = 0.002, d = 1.048$. In addition, participants were unable to explicitly match no-overlap endpoints shown in the same video during the final behavioral task (comparison to 0.2 chance level: $t(334) = -0.467, p = .641, BF_{01} = 4.141, d = -0.097$). In summary, participants rapidly learnt which scenes were from the same location; however, this was only seen in the overlap condition (and not in the no-overlap condition).

### Visual Representations of Specific Endpoints

First, we report the results of the mixed-effects model testing for representations of specific endpoints that remained relatively unchanged across sessions (i.e., pre-videos to post-videos). This revealed that correlations between the same endpoints (e.g., A1–A1, B1–B1) were greater than correlations between different endpoints (e.g., A1–A2, A1–B1) in both the right PHC, $t(13224) = 5.229, p_{FWE} < .001, BF_{01} = 0.001, d = 1.090$, and the left PHC, $t(13200) = 6.351, p_{FWE} < .001, BF_{01} < 0.001, d = 1.324$. This effect was not significant in either the right or left RSC, $t(13210) = 1.185, p_{FWE} = .945, BF_{01} = 2.454, d = 0.247$, and $t(13202) = -0.231, p_{FWE} = .999, BF_{01} = 4.463, d = -0.048$, respectively.

The searchlight analysis that tested for consistent representations of specific endpoints across the whole brain revealed representations in one large cluster that peaked in the right occipital lobe (area V1; $t(22) = 11.50, p_{FWE} < .001, k = 5202, BF_{01} < 0.001, d = 2.398$) and extended into the areas V2, V3, and V4 and the fusiform gyri bilaterally. Three smaller clusters were also detected in the right precuneus, $t(22) = 4.64, p_{FWE} = .011, k = 44, BF_{01} = 0.005, d = 0.968$, right inferior parietal lobule, $t(22) = 4.40, p_{FWE} = .028, k = 37, BF_{01} = 0.008, d = 0.918$, and right RSC, $t(22) = 4.32, p_{FWE} = .025, k = 38, BF_{01} = 0.008, d = 0.901$. The latter effect overlapped considerably with the right RSC ROI identified for each participant. However, the effect size estimated in the ROI analysis was weaker than the peak searchlight effect, principally because it was variable across endpoints and as such largely accounted for by random effects in the model. Unthresholded statistical maps of these effects are available at neurovault.org/collections/4819.

In summary, we find evidence that the PHC (bilaterally), the right RSC, and a number of early visual areas maintained consistent representations of specific endpoints across scanning sessions. Note that whether a region codes such representations across scanning sessions is independent of whether it may learn location-based memory representations in the second session; these effects are, in principle, dissociable.

As part of a supplementary analysis, we also tested for visual representations of specific scenes that remained stable within (but not necessarily across) scanning sessions (see osf.io/exzba/). To quantify the BOLD similarity of specific scenes within each session, we required two independent pattern representations per session. Thus, across both sessions, we estimated voxel patters derived from four distinct periods: (a) first half of Session 1 (pre-videos), (b) second half of Session 1 (pre-videos), (c) first half of Session 2 (post-videos), and (d) second half of Session 2 (post-videos). As a result, each of these voxel representations was only derived from four endpoint presentations. Nonetheless, when similarity scores were modeled in a mixed-effects regression, each ROI showed greater levels of similarity between representations of the same endpoint relative to the similarity between different endpoints (weakest effect in the left PHC: $t(26492) = 2.211$, $p = .027$, $BF_{01} = 0.606$, $d = 0.461$). Furthermore, this analysis revealed that representations of the same endpoints became more similar to one another after the videos in the right RSC and left PHC (weakest effect: $t(26492) = 2.598$, $p = .009$, $BF_{01} = 0.308$, $d = 0.542$). This latter effect was insensitive in the right PHC and left RSC (weakest effect: $t(26492) = 1.671$, $p = .095$, $BF_{01} = 1.375$, $d = 0.348$).

## Location-based Memory Representations

### Effects in the Right PHC

Next, we report the results of the mixed-effects model examining whether pattern similarity between different endpoints changed across sessions as a result of watching the videos. This revealed a significant Session × Condition interaction in the right PHC, $F(2, 2739) = 6.827$, $p_{FWE} = .004$ (average similarity matrices shown in Figure 3; condition estimates and confidence intervals plotted in Figure 4A). Post hoc tests showed that this effect was driven by a difference between pre-video to post-video sessions for endpoints in the overlap condition, $t(2739) = 2.923$, $p = .004$, $BF_{01} = 0.167$, $d = 0.610$. This difference was not observed in any other condition (no overlap: $t(2739) = 0.756$, $p = .450$, $BF_{01} = 3.533$, $d = 0.156$; unseen: $t(2739) = -0.970$, $p = .332$, $BF_{01} = 3.001$, $d = -0.202$). Furthermore, a significant Session × Overlap interaction highlighted that the similarity differences in the overlap condition were attributable to the video manipulation alone rather than some combination of other factors, $t(2739) = 2.549$, $p_{FWE} = .043$, $BF_{01} = 0.337$, $d = 0.532$.

Importantly, before the videos were shown, pairs of endpoints from the same location (i.e., A1–A2, B1–B2, and C1–C2) were found to evoke neural patterns that were more similar to each other than pairs of endpoints from different locations in the right PHC (e.g., A1–B2, B1–C2), $t(2739) = 2.498$, $p_{FWE} = .050$, $BF_{01} = 0.369$, $d = 0.521$ (see osf.io/uxhs9 for a plot of this effect). This "same-location" effect suggests that, even before the spatial relationship between scenes were known, the right PHC encoded visual properties of those scenes that generalized across different views. These data demonstrate that, despite controlling for similarity across stimuli using both the GIST descriptor and a pixel-wise correlation, and despite participants being unable to infer which endpoints were from the same location before watching the videos, we still found evidence for a "same-location" effect in the right PHC. This underlies the critical role of estimating pattern similarity before learning to identify significant increases in similarity post-video relative to pre-video (cf. Robertson et al., 2016). Note that this "same-location" effect is only seen when collapsing across



**Figure 3.** Mean representational similarity between endpoints in the right PHC, averaged across all participants and image sets. (A) Similarity between endpoints before the panoramic videos were shown (i.e., in Session 1). (B) Similarity between endpoints after the panoramic videos were shown (i.e., in Session 2). (C) Change in similarity that followed the panoramic videos (i.e., Session 2 minus Session 1). Color bars indicate both raw and baseline-adjusted Fisher $z$ statistics (above and below the color bar, respectively). Adjusted statistics account for trivial differences in similarity across scanning sessions caused by motion and scanner drift. This is achieved by subtracting out a baseline level of similarity between nonassociated endpoints (i.e., endpoints that were not from the same location, video, or experimental condition). Note that the baseline-adjusted statistics are shown for illustrative purposes only; each RSA was conducted on the raw Fisher $z$ statistics alone. Crosshatchings along the diagonal elements represent perfect correlations between identical BOLD responses and so were not included in the analyses.
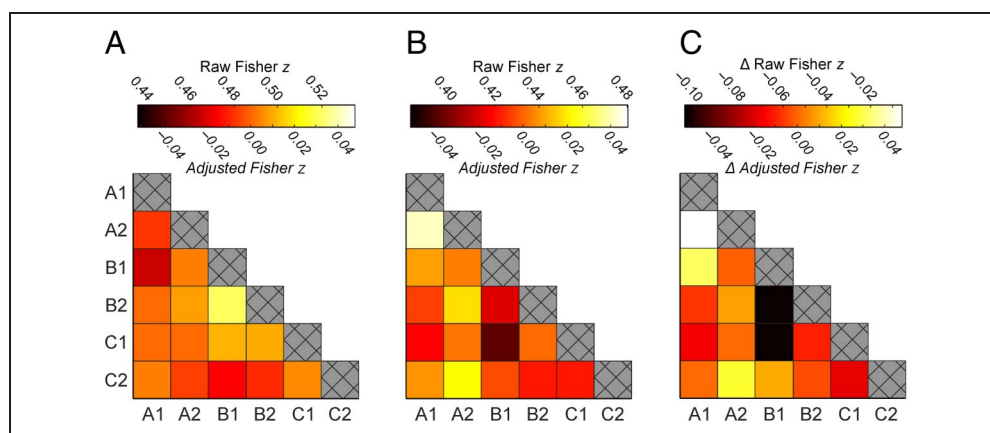
**Figure 4.** Results of the RSAs in the right PHC (rPHC; top row) and right RSC (rRSC; bottom row). (A) rPHC similarity estimates of scenes in the pre-video and post-video sessions, plotted by experimental condition. There was a significant change in similarity estimates between sessions in the overlap condition, $t(2739) = 2.923, p = .004, BF_{01} = 0.167, d = 0.610$, that was not present in the no-overlap and unseen conditions, $t(2739) = 0.756, p = .450, BF_{01} = 3.533, d = 0.158,$ and $t(2739) = -0.970, p = .332, BF_{01} = 3.001, d = -0.202$, respectively. (B) In the rPHC, pre-video to post-video changes in representational similarity for the overlap condition plotted against the number of correct matches between overlap endpoints in the post-video behavioral task. This association was not significant, $t(2739) = -0.892, p = .373, BF_{01} = 3.199, d = -0.186$. (C and D) Same as A and B but for the rRSC ROI. The rRSC showed no overall similarity changes in any of the experimental conditions ($t(2728) = 0.870, t(2728) = 1.419,$ and $t(2728) = -1.059$ for the overlap, no-overlap, and unseen conditions, respectively; all $ps > .156, BF_{01}s > 1.895, ds < 0.296$). Nonetheless, there was a significant association between behavioral performance and similarity changes in the overlap condition, $t(2728) = 2.886, p = .004, BF_{01} = 0.179, d = 0.602$. All bars plot baseline-corrected similarity estimates having subtracted out correlations between nonassociated endpoints (e.g., A1–B1, A1–B2). As such, the zero line in A and C denotes the average similarity of these nonassociated endpoints in each session. Error bars indicate 95% confidence intervals.
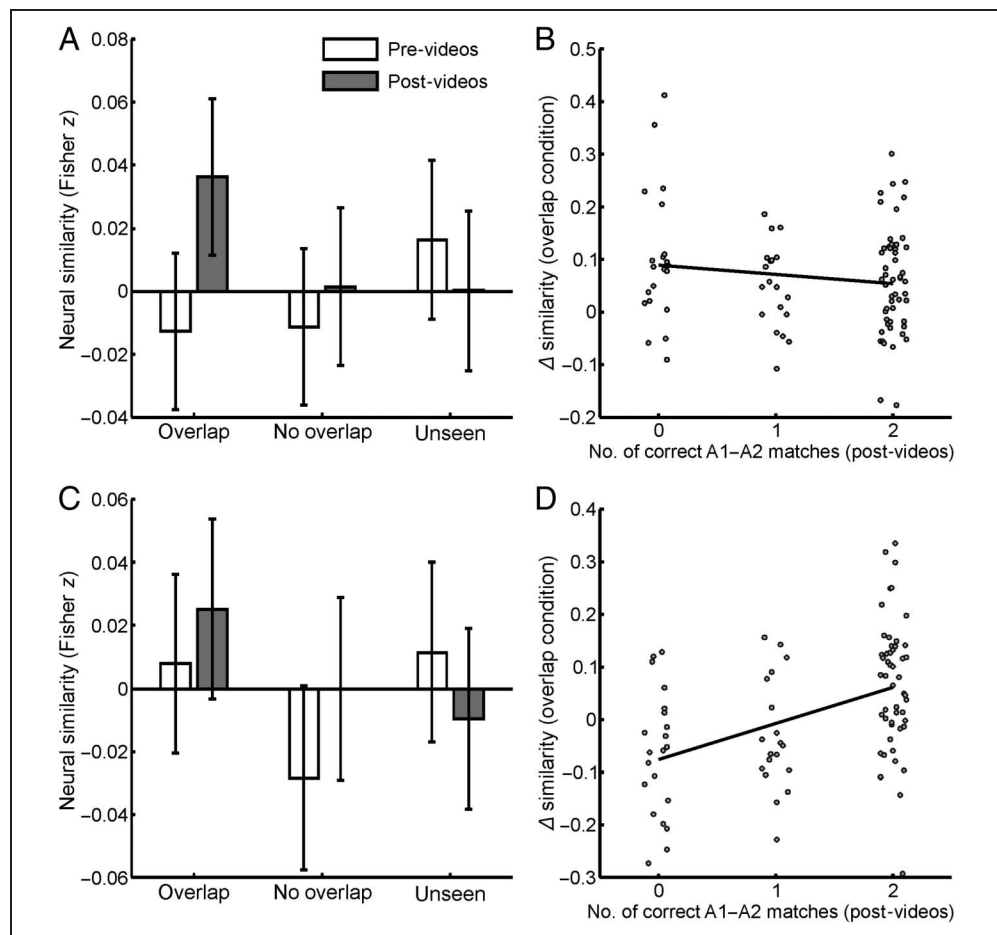
all endpoint pairs and is not evident in the Session 1 Overlap condition alone (osf.io/uxhs9).

### Effects in the Right RSC

The Session × Condition and Session × Overlap interactions were not significant in any other ROI ($Fs < 2.775, p_{FWE}s > .250$; similarity estimates for the right RSC plotted in Figure 4C). However, we saw a significant Session × Overlap × Behavior interaction in the right RSC, $t(2728) = 2.886, p_{FWE} = .016, BF_{01} = 0.179, d = 0.602$ (Figure 4D). This suggests that the RSC only encoded viewpoint-independent representations when the spatial relationships between endpoints could be retrieved during the postscanner test. No other ROIs showed a significant Session × Overlap × Behavior interaction (largest effect: $t = 0.050, p_{FWE} = 1, BF_{01} = 4.567, d = -0.010$).

### Differentiating the PHC and RSC

We next assessed whether there was evidence for dissociable roles of the right PHC and RSC, given that both represented location-based information but were differently associated with behavioral performance. Specifically, we assessed whether location-based representations in the RSC were significantly more associated with participants' ability to match endpoints from the same location compared to representations in the PHC. This would suggest that the RSC plays a greater role in guiding behavioral performance than the PHC. We therefore tested whether the Session × Overlap × Behavior (three-way) effect was larger in the RSC than the PHC. A comparison of effect sizes did show evidence for such a dissociation, $t(5311.9) = 3.931, p < .001, BF_{01} = 0.021, d = 0.820$.

This implies that the right PHC might have exhibited above-baseline pattern similarity between A1 and A2 endpoints even when those endpoints were not subsequently remembered as belonging to the same location. We directly tested this by rerunning the RSA having excluded A1–A2 pairs that were consistently remembered as belonging to the same location (i.e., having two correct responses during the postscanner test). Despite these exclusions, pattern similarity differences in the overlap condition remained significant, $t(1188) = 2.364, p = .018,$

$BF_{01} = 0.528$, $d = 0.493$, and were not seen in any other condition (no-overlap: $t(1188) = 0.324$, $p = .746$, $BF_{01} = 4.359$, $d = 0.068$; unseen: $t(1188) = -0.585$, $p = .559$, $BF_{01} = 3.915$, $d = -0.122$; see Figure 4B that plots the size of the Session × Overlap effect in the right PHC at each level of behavioral performance. In contrast, the right RSC only showed above-baseline pattern similarity when the endpoints were consistently remembered as belonging to the same location. Rerunning the RSA on these remembered pairs alone revealed similarity increases between consistently remembered endpoints in the overlap condition, $t(1538) = 2.449$, $p = .014$, $BF_{01} = 0.402$, $d = 0.511$ (see Figure 4D), that were not seen in any other condition (no-overlap: $t(1538) = 1.107$, $p = .269$, $BF_{01} = 2.651$, $d = 0.230$; unseen: $t(1538) = -1.316$, $p = .188$, $BF_{01} = 2.134$, $d = -0.274$).

In summary, we saw an increase in pattern similarity in the right PHC and right RSC between different scenes of the same location after they had been presented in an overlap video. Furthermore, we observed a dissociation between the PHC and the RSC. Whereas the PHC showed increased pattern similarity regardless of performance on the postscanner test, the RSC only showed increased pattern similarity when participants were able to subsequently identify those scenes as belonging to the same location.

### Across-Session Decreases in Pattern Similarity

Our mixed-effects regression models were conducted on the raw Fisher $z$ scores computed from each pair of endpoints. This ensured that effects were not driven by complex data manipulation or scaling, and so the data were not adjusted to account for across-session shifts in the similarity of all multivariate patterns (see Methods). Interestingly however, we did observe that Fisher $z$ scores decreased from pre-video to post-video across all pairs of endpoints regardless of condition, in each ROI (see figure at osf.io/2y3pm). This is reflected by a notable session effect in each mixed-effects model indicating reduced levels of similarity between nonassociated endpoints (i.e., endpoints not belonging to the same location, video, or experimental condition; minimum effect size: $t(2736) = -1.529$, $p = .126$, $BF_{01} = 1.655$, $d = 0.319$). As the size of this session effect was relatively large, the Session × Overlap and Session × Overlap × Behavior interactions involved less of a decrease in similarity scores relative to all other conditions (see Figure 3).

Given that similarity scores decrease across all endpoint pairs, it is unlikely that the Session effect was a direct result of our video manipulation (i.e., learning-induced neural differentiation). A mass differentiation on this scale would imply implausibly large amounts of information gain as the uniqueness (or entropy) of all neural representations would have to increase. Instead, it is more likely that the reduced levels of similarity were caused by systematically higher levels of noise in the second session. Most significantly, increases in temperature caused by radio frequency

absorption during scanning will shift the thermal equilibrium that governs how many hydrogen nuclei are aligned to the external magnetic field ($B_0$) and can therefore contribute to the MR signal (see osf.io/8kns6/). In this case, we would expect to see similar shifts in the level of similarity across the entire brain. To test this, we measured pattern similarity in the genu of the corpus callosum, a region that should exhibit negligible levels of BOLD activity. On the basis of a seed voxel at Montreal Neurological Institute of [0, 26, 6], multivariate patterns were taken from the 122 white-matter voxels closest to that seed in native space. The size of this ROI was chosen to reflect the average size of our a priori ROIs. A mixed-effects regression model of these data did indeed show reduced levels of neural similarity from Session 1 to Session 2, $t(2739) = -2.167$, $p = .030$, $BF_{01} = 0.651$, $d = -0.452$ (similar in magnitude to the session effect in all other regions; see osf.io/p9qzx/).

In summary, we conclude that the overall decrease in pattern similarity across sessions was not driven by any meaningful change in neural representations and, once controlled for, reveals a significant increase in pattern similarity in both the right PHC and RSC in the overlap condition, indicative of viewpoint-independent representations.

### Laterality of RSA Effects

The above analyses identified location-based representations in both right-hemisphere ROIs but no similar effects in the left hemisphere. Given this, we explored whether each RSA effect was significantly stronger in the right versus left hemisphere. Comparing the Session × Overlap effects in the PHC did indeed reveal a significantly stronger effect in the right hemisphere, $t(5390.5) = 3.798$, $p < .001$, $BF_{01} = 0.028$, $d = 0.792$. Similarly, comparing the Session × Overlap × Behavior interactions in the RSC revealed a significantly stronger effect in the right hemisphere, $t(5427.4) = 2.708$, $p = .007$, $BF_{01} = 0.251$, $d = 0.565$. Note that Robertson et al. (2016) collapsed their analyses across hemisphere, potentially masking laterality effects. These results are consistent with observations and theoretical models that the right hemisphere may preferentially process spatial information in humans as a consequence of predominantly left-lateralized language function (Shulman et al., 2010; Vallortigara & Rogers, 2005; Smith & Milner, 1981).

### Searchlight RSA

The searchlight analysis that tested for a Session × Overlap interaction across the whole brain revealed one small cluster in the right inferior occipital gyrus (Area V4), $t(21) = 4.78$, $p_{FWE} = .010$, $k = 38$, $BF_{01} < 0.003$, $d = 0.997$. However, when BOLD similarity in the cluster was modeled with the full mixed-effects analysis described above, the Session × Overlap effect was found to not be significant, $t(2740) = 1.734$, $p = .083$ (uncorrected), $BF_{01} = 1.259$, $d = 0.361$. Model parameter estimates suggested that the searchlight effect was driven by below-baseline BOLD similarity in the

overlap condition before the videos were shown (95% CI [−0.116, −0.021]), a result that is not consistent with any effect of interest. No other areas showed a significant Session × Overlap or Session × Overlap × Behavior interaction in the searchlight analysis. Nonetheless, both of the previously reported effects in the PHC and RSC are evident in the searchlight analysis at subthreshold levels, $t(21) > 2, d > 0.417$ (see neurovault.org/collections/4819/).

## Univariate Responses to Endpoints

We investigated whether each of our ROIs produced univariate BOLD activations consistent with a Session × Condition interaction or a three-way interaction with behavior. No such effects were found, all $Fs < 1.140$, $ps > .288$. Furthermore, a mass univariate analysis testing for these effects at the whole-brain level yielded no significant activations.

## Univariate Responses to Videos

Finally, we investigated whether univariate BOLD responses to the video clips differed between the overlap and no-overlap conditions or as a function of scene memory in the postscanner test. A group-level model was specified with predictors for (1) video type (overlap vs. no-overlap), (2) post-video performance in matching A1 and A2 endpoints, and (3) the interaction between video type and behavioral performance. This revealed two clusters that produced significantly greater BOLD responses during overlap versus no-overlap videos (Figure 5, hot colors). The largest of these peaked in the medial pFC and extended into the anterior cingulate, left frontal pole, and left middle frontal gyrus, $t(21) = 5.53, p_{FWE} < .001, k = 600, BF_{01} < 0.001, d = 1.153$. The second cluster peaked in the left supramarginal gyrus, $t(21) = 5.40, p_{FWE} = .004, k = 185, BF_{01} = 0.001, d = 1.126$, adjacent to a smaller, subthreshold effect in the left angular gyrus.

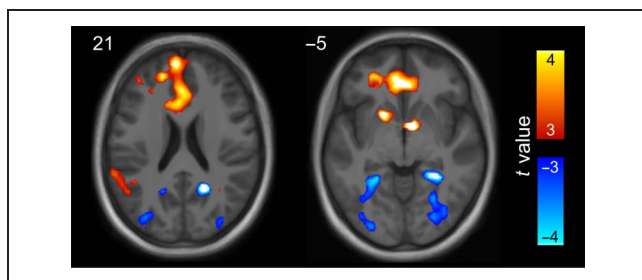No effects for the reverse contrast (i.e., no overlap > overlap) reached statistical significance at the whole-brain



**Figure 5.** Univariate BOLD effects showing differences in activity between the two video conditions (thresholded at $t(21) > 3, p < .004$ uncorrected, $BF_{01} < 0.143, d > 0.626$). Hot colors indicate areas showing a greater response to overlap versus no-overlap videos. Cool colors indicate areas showing a greater response to no-overlap versus overlap videos. An unthresholded statistical map of this contrast is available at neurovault.org/collections/4819.

level (subthreshold effects shown in Figure 5, cool colors). However, a small volume correction for the PHC and RSC bilaterally revealed two clusters with a significant no-overlap > overlap effect. These were found in the right RSC, $t(21) = −4.84, p_{FWE} = .032, k = 26, BF_{01} = 0.003, d = −1.001$, and right PHC, $t(21) = −4.77, p_{FWE} = .026, k = 30, BF_{01} = 0.003, d = −0.995$, extending into the fusiform gyrus. Subthreshold effects for the no-overlap > overlap contrast were also evident in the left RSC and PHC. These results were mirrored by a linear mixed-effects model contrasting overlap and no-overlap video responses averaged across each ROI in native space. Here, both the right PHC and right RSC exhibited greater BOLD activity in the no-overlap video condition relative to the overlap condition, $t(42) = −3.638, p_{FWE} = .003, BF_{01} = 0.039, d = −0.759$, and $t(42) = −3.499, p_{FWE} = .004, BF_{01} = 0.052, d = −0.730$, respectively. Effects in the left PHC and left RSC were below threshold and considerably weaker, $t(42) = −1.828, p_{FWE} = .299, BF_{01} = 1.101, d = −0.381$, and $t(42) = −2.212, p_{FWE} = .130, BF_{01} = 0.605, d = −0.461$, respectively. Neither the whole-brain analysis nor the mixed-effects model identified BOLD responses to the videos that significantly correlated with memory performance in the postscanner test.

In summary, we saw greater activity in the medial pFC during the overlap videos relative to the no-overlap videos. In contrast, the PHC and RSC showed greater activity during the no-overlap relative to overlap videos. In other words, the medial posterior regions that showed increased pattern similarity after presentation of the overlap video showed decreased activity while participants were watching the videos.

## DISCUSSION

We show that scene-selective brain regions rapidly learn location-based representations of novel environments by integrating information across different viewpoints. Once participants observed the spatial relationship between two viewpoints from a given location, BOLD pattern similarity between viewpoints increased in the right PHC and RSC, implying the emergence of location-based representations. In the right PHC, these representations appeared regardless of whether participants could identify which scenes were from the same location. In contrast, representations in the right RSC only emerged for scene pairs that participants could subsequently identify as being from the same location.

The results provide further evidence that the PHC and RSC support spatial representations that are not solely driven by visual features in a scene (Robertson et al., 2016; Marchette et al., 2015; Vass & Epstein, 2013; cf. Watson, Hartley, & Andrews, 2017). Using a similar panoramic video manipulation, Robertson et al. (2016) suggested that the RSC and OPA maintain viewpoint-independent representations but found a more general associative effect in the PHC. Our results further identify the PHC in this process

and highlight that RSC representations are more tightly linked to behavior. Note that the OPA was not one of our a priori ROIs, and we therefore make no claims in relation to this region supporting location-based representations (see ROIs section for further details). Our results also place constraints on models that describe how location-based representations are used. Unlike Robertson et al., we show that viewpoint-independent representations are evoked during passive viewing, in the absence of any explicit memory task (although we cannot rule out the possibility that participants engaged in active imagery, as explicitly required in Robertson et al.; see below).

Furthermore, we show that the learning of location-based representations can take place rapidly (in a single scanning session), with few exposures to the spatial layout of a location. Consistent with this, the firing fields of place cells have been shown to emerge rapidly in the rodent hippocampus (Monaco et al., 2014). Novel locations, where rats engaged in head-scanning behavior (i.e., exploration), were associated with place fields the next time the rat visited the same location. Our results provide evidence that location-based representations form after only three learning exposures to the videos. Although we were specifically interested in the emergence of viewpoint-independent spatial representations, a similar approach could be used to track the emergence of viewpoint-independent representations of other stimulus categories (e.g., objects or faces; see Clarke et al., 2016, for a similar approach), opening the door to understanding how such representations are formed, or modulated, across the visual system.

We also found that the right RSC only exhibited location-based representations when participants were able to identify which scenes belonged to that location in a postscanner test (PHC representations emerged regardless of behavioral performance on the postscanner test). This implies that the ability to identify differing scenes as from the same location is perhaps more dependent on representations in the RSC than PHC. Computational models hold that medial posterior and temporal regions (including the PHC and RSC) perform distinct but complementary functions in support of spatial navigation and imagery (Bicanski & Burgess, 2018; Byrne et al., 2007). Specifically, the PHC is thought to represent allocentric information related to the spatial geometry of the environment. Conversely, the posterior parietal cortex supports egocentric representations that allow the organism to actively navigate. The RSC transforms allocentric representations in the MTL into egocentric representations in the parietal cortex (and vice versa). Critically, the models predict that spatial navigation and planning is carried out in an egocentric reference frame. Thus, the RSC is critical to the translation of allocentric to more behaviorally relevant, egocentric information.

Our task required participants to match distinct scenes from the same location. This likely requires transformation from the presented egocentric viewpoint to an allocentric representation (egocentric-to-allocentric; i.e., A1 to the allocentric representation A*). In turn, the allocentric representation may allow for the retrieval of the associated viewpoint from the same location (allocentric-to-egocentric; i.e., A* to A2). Under this assumption, the RSC is likely to be more tightly coupled to behavior relative to the PHC, as shown in the present data. This is because allocentric representations in the PHC only require the initial egocentric-to-allocentric transformation to be retrieved (A1–A*). If only the egocentric-to-allocentric transformation occurs, participants will not be able to perform the task. As such, it is possible to see evidence for allocentric PHC representations in the absence of accurate behavior. For allocentric representations in the RSC to be retrieved, both the initial egocentric-to-allocentric (A1–A*) and subsequent allocentric-to-egocentric (A*–A2), transformation is required. If both transformations occur, then participants should be able to perform the task accurately. Thus, location-based representations in the RSC may only be seen in the presence of accurate behavior and may reflect the transformation between reference frames rather than reflecting an allocentric representation per se.

A related possibility is that, during the passive viewing of specific scenes, participants engaged in active imagery of the associated scenes, leading to subsequent improvements in behavior for scenes from the same location. However, we note that the task did not explicitly require memory retrieval; participants responded to oddball targets leaving little time for active imagery (see Linde-Domingo, Treder, Kerrén, & Wimber, 2019). In addition, participants would only be able to engage in active imagery on the overlap trials alone. Despite this, we did not observe any univariate BOLD effects indicative of additional processing on these trials. As such, the activation of these representations does not appear to depend on any task-specific memory demands. It is possible that the retrieval of PHC representations (i.e., egocentric-to-allocentric mapping) occurs relatively automatically, consistent with the proposal that allocentric representations in the MTL are automatically updated during self-motion in an environment (Bicanski & Burgess, 2018; Byrne et al., 2007). However, the retrieval of associated egocentric representations (i.e., allocentric-to-egocentric mapping) may not occur automatically during passive viewing, consistent with the observation that viewpoint-independent representations in the RSC are abolished when participants engage in a task that prevents them from active retrieval of spatial information (Marchette et al., 2015). Importantly, both of the above accounts are consistent with the proposal that the RSC plays a critical role in mapping between allocentric and egocentric representations.

Although consistent with models of allocentric processing, it is possible that the location-based representations we observed reflect other forms of associative learning (e.g., O'Reilly & Rudy, 2001). On this view, Scene A1 may become bound to A2 via a simple associative representation such that, after seeing the videos, A2 is covertly retrieved when presented with A1 (leading to increased

pattern similarity). However, contrary to our findings, this simple account may also predict increased similarity in the no-overlap condition, where B1 and C2 are shown in the same video—particularly given that models of associative learning often rely on prediction error signals to account for incidental encoding (Den Ouden, Friston, Daw, McIntosh, & Stephan, 2009), which could be strongest in the no-overlap condition. A second possibility is that the overlapping content in the overlap videos (relative to the no-overlap videos) increases the probability of a direct association between A1 and A2. Indeed, it is the overlapping content that likely drives the increase in pattern similarity between overlap endpoints. Our current study is not able to discern whether the resulting "location-based" representations are associative, or truly allocentric, in nature.

In terms of associative learning, a related possibility is that the overlapping content supports a more complex transitive representation (e.g., A1–AX and A2–AX where X is the overlapping scene in the center of the panorama). On this account, presentation of A1 cues the retrieval of AX and subsequently A2 (similar in nature to AB–AC inference paradigms; see Joensen, Gaskell, & Horner, 2020; Schlichting, Mumford, & Preston, 2015; Horner & Burgess, 2014; Schlichting, Zeithamova, & Preston, 2014; Zeithamova, Dominick, & Preston, 2012). Representations that encode these transitive relationships between scenes are possible and may support spatial navigation but are not directly predicted by models of spatial memory (Bicanski & Burgess, 2018; Byrne et al., 2007). Furthermore, the hippocampus and medial PFC (mPFC) are more typically associated with transitive inference (Schlichting et al., 2014, 2015; Zeithamova et al., 2012), yet we only found evidence of location-based representations in scene-selective regions. In addition, Robertson et al. have demonstrated that associative memory for scenes belonging to different locations is poor (comparable to their no-overlap condition) even when those scenes are presented in a "morphed" panorama such that they are associated with a common context. As such, our data are suggestive of processes that go beyond associative or transitive learning and provide support for models of allocentric processing, although we cannot rule out an "associative" explanation.

Finally, it is noteworthy that certain nonspatial models may be able to account for our findings. In particular, models of directed attention may predict increased levels of pattern similarity in the overlap condition if the overlap videos alerted participants to visual features that are shared across scenes (e.g., Luo, Roads, & Love, 2020; Mack, Preston, & Love, 2013). Further work will be needed to fully establish the true nature of the location-based representations that we report here. To fully match all visual features across scenes in each condition, one possibility would be to experimentally manipulate the central section of continuous panoramas so that no coherent spatial representation can be learned. Furthermore, to fully distinguish between allocentric and transitive (A1–AX–A2)

representations, an imaging study incorporating the panoramic morph manipulation used by Robertson et al. may be used.

Although we directly link to computational models of spatial navigation and imagery, as well as rodent studies on spatial navigation, it is important to note that we have assessed pattern similarity during visual presentation of static scenes. This is a common approach in human fMRI (Bonner & Epstein, 2017; Robertson et al., 2016; Marchette et al., 2015; Marchette et al., 2014), as it allows one to control for many potential experimental confounds that might be present in a more ecologically valid experimental setting (e.g., using virtual reality; Julian et al., 2016; Doeller, Barry, & Burgess, 2010). However, this approach has the issue of being further removed from real-world spatial navigation. Interestingly, we saw evidence for increases in pattern similarity despite using a low-level attentional task, speaking to the potential automaticity of retrieving more location-based representations. Across the literature, there are numerous examples of evidence for spatial learning in humans and rodents during goal-directed navigation (Aoki, Igata, Ikegaya, & Sasaki, 2019; Howard et al., 2014), non-goal-directed navigation (e.g., O'Keefe & Dostrovsky, 1971; Tolman, 1948), mental imagery (e.g., Bellmund, Deuker, Schröder, & Doeller, 2016; Horner, Bisby, Zotow, Bush, & Burgess, 2016), and viewing of static images (e.g., Robertson et al., 2016; Marchette et al., 2015; Vass & Epstein, 2013). Our study adds to this growing literature suggesting that these representations can be assessed across diverse experimental environments with multiple methodologies.

The PHC has been proposed to represent several complementary spatial representations, including geometric information regarding one's location in relation to bearings and distances to environmental features (e.g., boundaries; Park, Brady, Greene, & Oliva, 2011). The representations that we observed in PHC may reflect enriched spatial representations relating specific scenes to environmental features outside the current field of view. Also consistent with our results, the PHC may represent spatial contexts more broadly (Epstein & Vass, 2014). The experimental manipulation used here could be modified to learn novel locations in the same spatial context, potentially dissociating between the above accounts. A further proposal is that viewpoint-independent representations in the PHC reflect prominent landmarks that are visible across viewpoints (Marchette et al., 2015). Although this proposal yields similar predictions to above, it is less able to account for our finding of shared representations of views that did not contain any of the same landmarks.

Our PHC results are somewhat inconsistent with those of Robertson et al. (2016). Whereas our similarity increases were specific to the overlap condition, Robertson et al. saw effects in both their overlap and no-overlap conditions. One possibility is that our results reflect a Type II error, in that we failed to find an effect in the no-overlap

condition when one is present. A second possibility is that Robertson et al. either (1) found an effect in the no-overlap condition when one is not present (i.e., a Type I error) or (2) failed to find a similarity effect in the overlap condition that was significantly larger than in the no-overlap condition (a Type II error). Notably, the "overlap > no-overlap" effect size that we observed in the PHC is considerably larger than the same effect reported by Robertson et al. (0.610 relative to −0.062) and more in line with their RSC and OPA effects (0.470 and 0.415, respectively). Thus, it seems plausible that the disagreement stems from a Type I error in Robertson et al. Despite this, without further information, it is not possible to draw clear conclusions.

However, one important caveat is that we also saw evidence for a "same-location" effect in the PHC before learning had occurred. This effect was seen despite controlling for visual similarity across stimuli using the GIST descriptor, accounting for pixel-wise correlations in luminance and color content, and despite participants being unable to identify which endpoints were from the same location before the videos. It is therefore possible that the PHC effects in Robertson et al. could have been driven by a similar effect not dependent on learning. This underlines the importance of including a prelearning versus postlearning estimate of pattern similarity, to definitively rule out trivial effects driven by preexisting similarities between images that are difficult to control for.

RSC representations may reflect the retrieval of spatial or conceptual information associated with the environment (Marchette et al., 2015). Further evidence suggests that the RSC contains multiple viewpoint-dependent and viewpoint-independent (Vass & Epstein, 2013), as well as local and global (Jacob et al., 2017; Marchette et al., 2014), spatial representations. This multitude of representations fits with the proposed role of the RSC as a transformation circuit, mapping between allocentric and egocentric representations. The heterogeneity of representations, relative to the PHC, may also be a further reason why we did not see clear evidence for location-based representations without taking behavior into account. Our RSC results are consistent with those of Robertson et al. in that they saw a clear effect after more extensive learning across 2 days (where behavioral performance was likely higher than in our study). However, we extend these findings to show that these effects are specifically associated with the locations that each individual participant has learned (i.e., a within-participant correlation that is consistent across participants). Regardless of the exact nature of such representations, our results provide clear evidence that we can track their emergence in both the PHC and RSC.

Although more explorative, we also examined activity during learning of new spatial relationships (i.e., video presentation). BOLD activations in medial posterior brain regions (including but not limited to the PHC and RSC ROIs) were greater for no-overlap videos compared to overlap videos. This effect perhaps reflects greater fMRI adaptation during the overlap videos because they presented the central viewpoint of the panorama more frequently than no-overlap videos (Figure 1). However, it is interesting that the same cortical regions that showed increased pattern similarity after presentation of the overlap video showed decreased activity when participants were watching the videos. This underlines the complex relationship between univariate activity during learning and resultant changes in patterns of activity after learning. More theoretically driven research would be needed to provide a robust explanation for this finding.

In addition, we found that mPFC showed a greater BOLD response in the overlap than no-overlap condition. This may reflect a mnemonic integration process that guides the learning of viewpoint-independent representations. Similar effects in mPFC have been observed in tasks that require integrating overlapping memories to support inference and generalization (Milivojevic, Vicente-Grabovetsky, & Doeller, 2015; Schlichting et al., 2015). Indeed, mPFC has been implicated in detecting new information that is congruent with previously learnt materials so that it can be integrated into a generalized representation (van Kesteren, Ruiter, Fernández, & Henson, 2012). Our results are broadly in line with this proposal, where mPFC may be detecting the presence of overlapping spatial information during the overlap videos, resulting in the integration of previously learnt representations into more coherent viewpoint-independent representations in posterior medial regions. Despite this, our results do not exclude the possibility that mPFC activations reflect disinhibition from medial–posterior inputs (which showed reduced activity) or attentional differences related to the behavioral task.

We have shown that brain regions in the scene network, specifically the right PHC and RSC, rapidly learn representations of novel environments by integrating information across different viewpoints. They appear to be relatively viewpoint-independent in that they become active regardless of which part of an environment is in the current field of view. We show that the PHC and RSC have potentially dissociable roles, consistent with models that propose the RSC plays a role in translating viewpoint-independent representations into a behaviorally relevant egocentric reference frame. Finally, our experimental approach allows for tracking the emergence of viewpoint-independent representations across the visual system.

## Author Contributions

## REFERENCES

Andersson, J. L. R., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling geometric deformations in EPI time series. *Neuroimage*, *13*, 903–919. **DOI:** https://doi.org/10.1006/nimg.2001.0746, **PMID:** 11304086

Aoki, Y., Igata, H., Ikegaya, Y., & Sasaki, T. (2019). The integration of goal-directed signals onto spatial maps of hippocampal place cells. *Cell Reports*, *27*, 1516–1527. **DOI:** https://doi.org/10.1016/j.celrep.2019.04.002, **PMID:** 31042477

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, *38*, 95–113. **DOI:** https://doi.org/10.1016/j.neuroimage.2007.07.007, **PMID:** 17761438

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. **DOI:** https://doi.org/10.1016/j.jml.2007.12.005

Bellmund, J. L. S., Deuker, L., Schröder, T. N., & Doeller, C. F. (2016). Grid-cell representations in mental simulation. *eLife*, *5*, e17089. **DOI:** https://doi.org/10.7554/eLife.17089, **PMID:** 27572056, **PMCID:** PMC5005038

Bicanski, A., & Burgess, N. (2018). A neural-level model of spatial memory and imagery. *eLife*, *7*, e33752. **DOI:** https://doi.org/10.7554/eLife.33752, **PMID:** 30176988, **PMCID:** PMC6122954

Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences, U.S.A.*, *114*, 4793–4798. **DOI:** https://doi.org/10.1073/pnas.1618228114, **PMID:** 28416669, **PMCID:** PMC5422815

Burgess, N., Becker, S., King, J. A., & O'Keefe, J. (2001). Memory for events and their spatial context: Models and experiments. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *356*, 1493–1503. **DOI:** https://doi.org/10.1098/rstb.2001.0948, **PMID:** 11571039, **PMCID:** PMC1088531

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*, 340–375. **DOI:** https://doi.org/10.1037/0033-295X.114.2.340, **PMID:** 17500630, **PMCID:** PMC2678675

Calton, J. L., & Taube, J. S. (2009). Where am I and how will I get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. *Neurobiology of Learning and Memory*, *91*, 186–196. **DOI:** https://doi.org/10.1016/j.nlm.2008.09.015, **PMID:** 18929674, **PMCID:** PMC2666283

Clarke, A., Pell, P. J., Ranganath, C., & Tyler, L. K. (2016). Learning warps object representations in the ventral temporal cortex. *Journal of Cognitive Neuroscience*, *28*, 1010–1023. **DOI:** https://doi.org/10.1162/jocn_a_00951, **PMID:** 26967942

Den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex*, *19*, 1175–1185. **DOI:** https://doi.org/10.1093/cercor/bhn161, **PMID:** 18820290, **PMCID:** PMC2665159

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*, 657–661. **DOI:** https://doi.org/10.1038/nature08704, **PMID:** 20090680, **PMCID:** PMC3173857

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*, 109–120. **DOI:** https://doi.org/10.1016/j.neuron.2004.08.028, **PMID:** 15450164

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, *20*, 1504–1513. **DOI:** https://doi.org/10.1038/nn.4656, **PMID:** 29073650, **PMCID:** PMC6028313

Epstein, R. A., & Vass, L. K. (2014). Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *369*, 20120533. **DOI:** https://doi.org/10.1098/rstb.2012.0533, **PMID:** 24366141, **PMCID:** PMC3866451

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, *2*, 1360–1383. **DOI:** https://doi.org/10.1214/08-AOAS191

Hannula, D. E., & Ranganath, C. (2009). The eyes have it: Hippocampal activity predicts expression of memory in eye movements. *Neuron*, *63*, 592–599. **DOI:** https://doi.org/10.1016/j.neuron.2009.08.025, **PMID:** 19755103, **PMCID:** PMC2747814

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*, 1315–1326. **DOI:** https://doi.org/10.1002/hipo.20857, **PMID:** 20928831

Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like processing of imagined navigation. *Current Biology*, *26*, 842–847. **DOI:** https://doi.org/10.1016/j.cub.2016.01.042, **PMID:** 26972318, **PMCID:** PMC4819517

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology*, *24*, 988–992. **DOI:** https://doi.org/10.1016/j.cub.2014.03.012, **PMID:** 24746796, **PMCID:** PMC4012134

Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., et al. (2014). The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. *Current Biology*, *24*, 1331–1340. **DOI:** https://doi.org/10.1016/j.cub.2014.05.001, **PMID:** 24909328, **PMCID:** PMC4062938

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., & Turner, R. (2002). Image distortion correction in fMRI: A quantitative evaluation. *Neuroimage*, *16*, 217–240. **DOI:** https://doi.org/10.1006/nimg.2001.1054, **PMID:** 11969330

Jacob, P.-Y., Casali, G., Spieser, L., Page, H., Overington, D., & Jeffery, K. (2017). An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. *Nature Neuroscience*, *20*, 173–175. **DOI:** https://doi.org/10.1038/nn.4465, **PMID:** 27991898, **PMCID:** PMC5274535

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of Experimental Psychology: General*, *149*, 230–248. **DOI:**

https://doi.org/10.1037/xge0000648, **PMID:** 31305093, **PMCID:** PMC6951107

Julian, J. B., 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. **DOI:** https://doi.org/10.1016/j.neuroimage.2012.02.055, **PMID:** 22398396

Julian, J. B., Keinath, A. T., Marchette, S. A., & Epstein, R. A. (2018). The neurocognitive basis of spatial reorientation. *Current Biology*, *28*, R1059–R1073. **DOI:** https://doi.org/10.1016/j.cub.2018.04.057, **PMID:** 30205055, **PMCID:** PMC6161705

Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The occipital place area is causally involved in representing environmental boundaries during navigation. *Current Biology*, *26*, 1104–1109. **DOI:** https://doi.org/10.1016/j.cub.2016.02.066, **PMID:** 27020742, **PMCID:** PMC5565511

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795. **DOI:** https://doi.org/10.1080/01621459.1995.10476572

Kumaran, D., Hassabis, D., Spiers, H. J., Vann, S. D., Vargha-Khadem, F., & Maguire, E. A. (2007). Impaired spatial and non-spatial configural learning in patients with hippocampal pathology. *Neuropsychologia*, *45*, 2699–2711. **DOI:** https://doi.org/10.1016/j.neuropsychologia.2007.04.007, **PMID:** 17507060, **PMCID:** PMC2572909

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*, 179. **DOI:** https://doi.org/10.1038/s41467-018-08080-2, **PMID:** 30643124, **PMCID:** PMC6331625

Luo, X., Roads, B. D., & Love, B. C. (2020). The costs and benefits of goal-directed attention in deep convolutional neural networks. *ArXiv Preprint arXiv:2002.02342*. Retrieved from http://arxiv.org/abs/2002.02342.

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology*, *23*, 2023–2027. **DOI:** https://doi.org/10.1016/j.cub.2013.08.035, **PMID:** 24094852, **PMCID:** PMC3874407

Malcolm, G. L., Silson, E. H., Henry, J. R., & Baker, C. I. (2018). Transcranial magnetic stimulation to the occipital place area biases gaze during scene viewing. *Frontiers in Human Neuroscience*, *12*, 189. **DOI:** https://doi.org/10.3389/fnhum.2018.00189, **PMID:** 29867413, **PMCID:** PMC5953332

Marchette, S. A., Ryan, J., & Epstein, R. A. (2017). Schematic representations of local environmental space guide goal-directed navigation. *Cognition*, *158*, 68–80. **DOI:** https://doi.org/10.1016/j.cognition.2016.10.005, **PMID:** 27814459, **PMCID:** PMC5123926

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2014). Anchoring the neural compass: Coding of local spatial reference frames in human medial parietal lobe. *Nature Neuroscience*, *17*, 1598–1606. **DOI:** https://doi.org/10.1038/nn.3834, **PMID:** 25282616, **PMCID:** PMC4309016

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside looking in: Landmark generalization in the human navigational system. *Journal of Neuroscience*, *35*, 14896–14908. **DOI:** https://doi.org/10.1523/JNEUROSCI.2270-15.2015, **PMID:** 26538658, **PMCID:** PMC4635136

Milivojevic, B., Vicente-Grabovetsky, A., & Doeller, C. F. (2015). Insight reconfigures hippocampal–prefrontal memories. *Current Biology*, *25*, 821–830. **DOI:** https://doi.org/10.1016/j.cub.2015.01.033, **PMID:** 25728693

Monaco, J. D., Rao, G., Roth, E. D., & Knierim, J. J. (2014). Attentive scanning behavior drives one-trial potentiation of hippocampal place fields. *Nature Neuroscience*, *17*, 725–731.

**DOI:** https://doi.org/10.1038/nn.3687, **PMID:** 24686786, **PMCID:** PMC4036486

Motley, S. E., Grossman, Y. S., Janssen, W. G. M., Baxter, M. G., Rapp, P. R., Dumitriu, D., et al. (2018). Selective loss of thin spines in area 7a of the primate intraparietal sulcus predicts age-related working memory impairment. *Journal of Neuroscience*, *38*, 10467–10478. **DOI:** https://doi.org/10.1523/jneurosci.1234-18.2018, **PMID:** 30355632, **PMCID:** PMC6284109

O'Keefe, J., & Burgess, N. (2005). Dual phase and rate coding in hippocampal place cells: Theoretical significance and relationship to entorhinal grid cells. *Hippocampus*, *15*, 853–866. **DOI:** https://doi.org/10.1002/hipo.20115, **PMID:** 16145693, **PMCID:** PMC2677681

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*, 171–175. **DOI:** https://doi.org/10.1016/0006-8993(71)90358-1, **PMID:** 5124915

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175. **DOI:** https://doi.org/10.1023/A:1011139631724

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*, 311–345. **DOI:** https://doi.org/10.1037/0033-295x.108.2.311, **PMID:** 11381832

Park, S., Brady, T. F., Greene, M. R., & 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. *Journal of Neuroscience*, *31*, 1333–1340. **DOI:** https://doi.org/10.1523/JNEUROSCI.3885-10.2011, **PMID:** 21273418, **PMCID:** PMC6623596

Robertson, C. E., Hermann, K. L., Mynick, A., Kravitz, D. J., & Kanwisher, N. (2016). Neural representations integrate the current field of view with the remembered 360° panorama in scene-selective cortex. *Current Biology*, *26*, 2463–2468. **DOI:** https://doi.org/10.1016/j.cub.2016.07.002, **PMID:** 27618266

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 8151. **DOI:** https://doi.org/10.1038/ncomms9151, **PMID:** 26303198, **PMCID:** PMC4560815

Schlichting, M. L., Zeithamova, D., & Preston, A. R. (2014). CA1 subfield contributions to memory integration and inference. *Hippocampus*, *24*, 1248–1260. **DOI:** https://doi.org/10.1002/hipo.22310, **PMID:** 24888442, **PMCID:** PMC4159432

Shulman, G. L., Pope, D. L. W., Astafiev, S. V., McAvoy, M. P., Snyder, A. Z., & Corbetta, M. (2010). Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *Journal of Neuroscience*, *30*, 3640–3651. **DOI:** https://doi.org/10.1523/jneurosci.4085-09.2010, **PMID:** 20219998, **PMCID:** PMC2872555

Silson, E. H., Steel, A. D., & Baker, C. I. (2016). Scene selectivity and retinotopy in medial parietal cortex. *Frontiers in Human Neuroscience*, *10*, 412. **DOI:** https://doi.org/10.3389/fnhum.2016.00412, **PMID:** 27588001, **PMCID:** PMC4988988

Smith, M. L., & Milner, B. (1981). The role of the right hippocampus in the recall of spatial location. *Neuropsychologia*, *19*, 781–793. **DOI:** https://doi.org/10.1016/0028-3932(81)90090-7, **PMID:** 7329524

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208. **DOI:** https://doi.org/10.1037/h0061626, **PMID:** 18870876

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical

labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–289. **DOI:** https://doi.org/10.1006/nimg.2001.0978, **PMID:** 11771995

Vallortigara, G., & Rogers, L. J. (2005). Survival with an asymmetrical brain: Advantages and disadvantages of cerebral lateralization. *Behavioral and Brain Sciences*, *28*, 575–589. **DOI:** https://doi.org/10.1017/s0140525x05000105, **PMID:** 16209828

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*, 211–219. **DOI:** https://doi.org/10.1016/j.tins.2012.02.001, **PMID:** 22398180

Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *Journal of Neuroscience*, *33*, 6133–6142. **DOI:** https://doi.org/10.1523 /JNEUROSCI.3873-12.2013, **PMID:** 23554494, **PMCID:** PMC3656495

Watson, D. M., Hartley, T., & Andrews, T. J. (2017). Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex. *Cortex*, *92*, 162–174. **DOI:** https://doi.org/10.1016/j.cortex .2017.04.011, **PMID:** 28499144

Welch, B. L. (1947). The generalization of "student's" problem when several different population variances are involved. *Biometrika*, *34*, 28–35. **DOI:** https://doi.org/10.1093/biomet /34.1-2.28, **PMID:** 20287819

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*, 168–179. **DOI:** https://doi.org/10.1016/j.neuron.2012 .05.010, **PMID:** 22794270, **PMCID:** PMC3398403