Behavioral/Cognitive

Predictive Representations in Hippocampal and Prefrontal Hierarchies

Iva K. Brunec[1] and Ida Momennejad[2]

[1]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, and [2]Microsoft Research, New York, New York 10012

As we navigate the world, we use learned representations of relational structures to explore and to reach goals. Studies of how relational knowledge enables inference and planning are typically conducted in controlled small-scale settings. It remains unclear, however, how people use stored knowledge in continuously unfolding navigation (e.g., walking long distances in a city). We hypothesized that multiscale predictive representations guide naturalistic navigation in humans, and these scales are organized along posterior-anterior prefrontal and hippocampal hierarchies. We conducted model-based representational similarity analyses of neuroimaging data collected while male and female participants navigated realistically long paths in virtual reality. We tested the pattern similarity of each point, along each path, to a weighted sum of its successor points within predictive horizons of different scales. We found that anterior PFC showed the largest predictive horizons, posterior hippocampus the smallest, with the anterior hippocampus and orbitofrontal regions in between. Our findings offer novel insights into how cognitive maps support hierarchical planning at multiple scales.

Key words: cognitive maps; hippocampus; navigation; predictive representations; PFC; successor representation

---

Significance Statement

Whenever we navigate the world, we represent our journey at multiple horizons: from our immediate surroundings to our distal goal. How are such cognitive maps at different horizons simultaneously represented in the brain? Here, we applied a reinforcement learning-based analysis to neuroimaging data acquired while participants virtually navigated their hometown. We investigated neural patterns in the hippocampus and PFC, key cognitive map regions. We uncovered predictive representations with multiscale horizons in prefrontal and hippocampal gradients, with the longest predictive horizons in anterior PFC and the shortest in posterior hippocampus. These findings provide empirical support for the computational hypothesis that multiscale neural representations guide goal-directed navigation. This advances our understanding of hierarchical planning in everyday navigation of realistic distances.

Introduction

When we navigate a city, we draw on our memory. We learn, retrieve, and update representations of relationships among

different locations. This relational knowledge guides decisions and behavior (O'Keefe & Nadel, 1978; Burgess et al. 2002; Behrens et al., 2018; Momennejad, 2020) and has been captured by computational models of planning, inference, and spatial navigation (Garvert et al., 2017; Momennejad et al., 2017; Stachenfeld et al., 2017). While some argue for one-step relational representations, here we scale evidence for computational models suggesting that predictive representations of relational structures may be organized at multiple scales along hippocampal (Stachenfeld et al., 2017; Momennejad and Howard, 2018) and prefrontal (Christoff and Gabrieli, 2000; Koechlin and Hyafil, 2007; Momennejad and Haynes, 2013) hierarchies. Such hierarchical structure could also enable the extraction of abstract relational structures that unfold at lower levels (Fig. 1A).

Previously, we had shown that representations learned by reinforcement learning (RL) models capture human planning (Momennejad et al., 2017), in highly controlled experiments with fixed predictive scales. Here, we used model-based fMRI data analysis to test predictions of a multiscale predictive
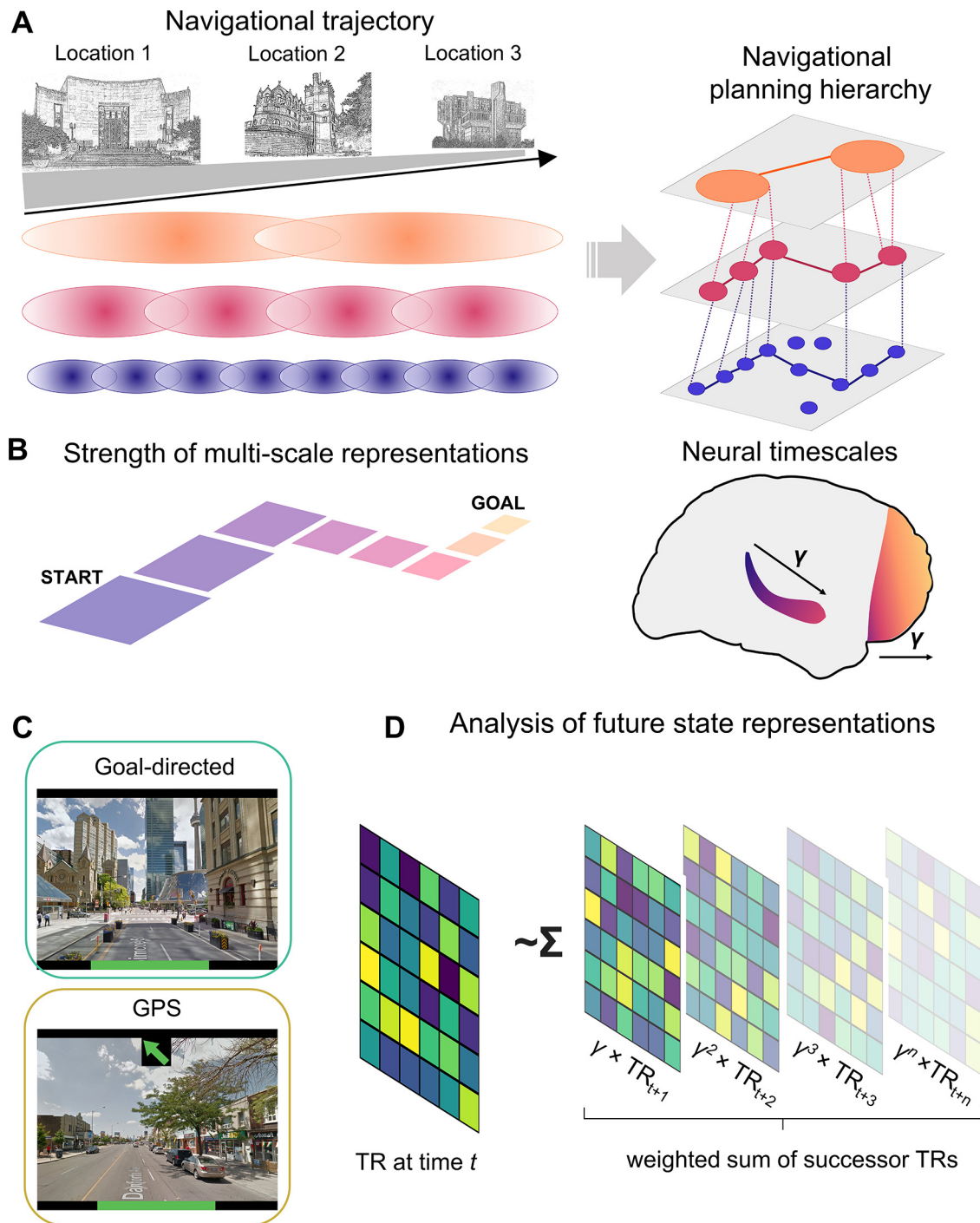
Figure 1. Schematic of the hypothesis, task conditions, and analytic methods. **A**, Multiple scales of representation along a navigated route are activated simultaneously. Longer predictive horizons correspond to longer-range planning and greater scales of navigational representations. **B**, Predictive representations in the hippocampus and PFC should proceed along a posterior-anterior gradient within the hippocampus and PFC. **C**, Participants used the same keys to navigate goal-directed routes and to follow the GPS dynamic arrow, but only goal-directed routes required goal-directed navigation. **D**, Analytic approach. The voxelwise pattern at each time point was correlated with the γ-weighted sum of all future states (for γ values of 0.1, 0.6, 0.8, and 0.9). With higher γ values, the weighted future states remain >0 further into the future. Not displayed: We also computed similarity for each TR to goal, and similarity of each TR to mean of future TRs (equally weighted) within a given horizon (e.g., 10 TRs). For details of the model-based analysis, see Materials and Methods.

representation model (Momennejad and Howard, 2018) on brain signals collected during virtual navigation of a real-world city that participants had learned in their daily lives (dataset from Brunec et al., 2018). Our results show that representations in prefrontal and hippocampal hierarchies during real-world navigation were aligned with model predictions.

Our framework of multiscale representations stems as much from computational models as decades of electrophysiology and

neuroimaging findings. Our first hypothesis was that, during virtual navigation, anterior hippocampus would display representational similarity at longer predictive scales than posterior hippocampus. Rodent place field size increases along the dorsal-ventral hippocampal axis, with larger and more overlapping place fields in more ventral regions (Jung et al., 1994; Kjelstrup et al., 2008; Strange et al., 2014; Contreras et al., 2018). Human fMRI evidence suggests a similar gradient along the posterior-

anterior axis (homologous to the rodent dorsal-ventral axis) (Poppenk et al., 2013). The larger-scale anterior hippocampal representations might support goal-directed search (Ruediger et al., 2012), integration of spatial and nonspatial states further apart (Collin et al., 2015), and longer time horizons (Nielson et al., 2015). Posterior hippocampal representations are more myopic and may support fine-grained spatial relations (Evensmoen et al., 2013) and pattern separation in memory (Schlichting et al., 2015; Duncan and Schlichting, 2018; Lohnas et al., 2018). Recent computational models of predictive representations capture multiscale place fields and why they skew toward goals (Stachenfeld et al., 2017; Momennejad and Howard, 2018).

Our second hypothesis was that anterior PFC (antPFC) would display representational similarity to more distant states than posterior PFC. The PFC's hierarchical representations (Badre and D'Esposito, 2007) support active navigation and planning (Spiers and Gilbert, 2015; Epstein et al., 2017), computing alternative paths to goal (Javadi et al., 2017), reversals and detours (Spiers and Gilbert, 2015), and retrospective revaluation via offline replay (Momennejad et al., 2018). Neuroimaging evidence suggests a prefrontal hierarchy whereby more antPFC regions support relational reasoning (Christoff et al., 2009), abstraction (Christoff et al., 2001; Bunge et al., 2003), and prospective memory (Gilbert, 2011; Momennejad and Haynes, 2012, 2013).

Finally, we hypothesized that representational scales in antPFC would exceed the longest predictive horizons of hippocampal representations (Fig. 1B). The antPFC is the largest cytoarchitectonic region of the human PFC (Ramnani and Owen, 2004). PFC's recurrent interconnectivity enables information to linger across longer scales allowing slower learning and integration. In contrast, the hippocampus supports rapid statistical learning (Schapiro et al., 2013, 2016, 2017) and is less heterogeneous across mammalian species (Strange et al., 2014).

To test these hypotheses, we conducted model-based representational similarity analyses (Fig. 1D) on an existing fMRI dataset (Brunec et al., 2018), in which participants actively navigated to known goals (goal-directed condition), or followed a dynamic arrow along unfamiliar routes (GPS condition) in a virtual version of Toronto (Fig. 1C). The participants' experience in this virtual setup was as realistic as possible within the constraints of fMRI, and benefited from real-world familiarity, allowing us to compare predictive horizons on well-learned versus novel routes.

Consistent with our predictions, antPFC displayed representational similarity at longer horizons on goal-directed compared with GPS-guided paths. Anterior hippocampus followed the PFC, whereas posterior hippocampus supported smallest-scale predictive representations.

Materials and Methods

Subjects. Twenty-two healthy right-handed volunteers were recruited. One participant was excluded because of excessive difficulty with the task (i.e., repeatedly getting lost). Two additional participants were excluded because of incomplete data or technical issues. Exclusions resulted in 19 participants who completed the study (9 males; mean age 22.58 years, range 19-30 years). The sample size was not predetermined using a power analysis. We reanalyzed the dataset from a previously published study and included all participants with usable data. All participants had lived in Toronto for at least 2 years (mean = 10.45, SE = 1.81). All participants were free of psychiatric and neurologic conditions, had normal or corrected-to-normal vision, and otherwise met the criteria for participation in fMRI studies. Informed consent was obtained from all participants in accordance with Rotman Research Institute at Baycrest's

ethical guidelines. Participants received monetary compensation on completion of the study.

Experimental design and paradigm. The details of the experimental design have been reported previously (Brunec et al., 2018). The task used a realistic navigation software drawing on 360° panoramic images from Google Street View. This allowed participants to walk through a virtual Toronto from a first-person, street-level perspective. The navigation software was written in MATLAB version 7.5.0.342. Navigation was controlled using three buttons: left, right, and forward. A "done" button allowed participants to indicate that they had completed a route. The task was projected on a screen in the bore of the scanner viewed by the participants through a mirror mounted inside of the head coil. Participants navigated in four conditions, and navigated 16 routes in total (four in each condition, in a randomized order).

Data from two conditions of interest were analyzed in the present manuscript: goal-directed and GPS/arrow-following routes. The routes were constructed before the day of scanning: participants built routes with researcher assistance, using a computer program which showed overhead maps of Toronto. Additionally, sets of routes in areas of Toronto with which participants were generally unfamiliar were created. Four of these routes were randomly assigned to each participant to be used in the baseline (GPS) condition. In the scanner, participants were provided with goal-directed route destinations and asked to navigate toward the goal along the most goal-directed/comfortable route. GPS trials involved no goal-directed navigation; instead, participants followed a dynamic arrow (Fig. 1C). To navigate GPS-guided routes, participants used the same control buttons as they did along goal-directed routes. However, in the GPS condition, they did not know the goal or the distance. We only analyzed routes where participants successfully reached the goal (M-$_{goal-directed}$ = 3.37, M-$_{GPS}$ = 3.16 routes). Comparing these conditions enabled us to contrast navigational signals associated with goal-directed navigation with matched motor control and optic flow, but no goal.

fMRI acquisition and preprocessing. Participants were scanned with a 3T Siemens MRI scanner at the Rotman Research Institute at Baycrest. A high-resolution 3D MPRAGE T1-weighted pulse sequence image (160 axial slices, 1 mm thick, FOV = 256 mm) was first obtained to register functional maps against brain anatomy. Functional T2*-weighted images were acquired using EPI (30 axial slices, 5 mm thick, TR = 2000 ms, TE = 30 ms, flip angle = 70 degrees, FOV = 200 mm). The native EPI resolution was 64 × 64 with a voxel size of 3.5 mm × 3.5 mm × 5.0 mm. The preprocessed data were the same as those used in Brunec et al., (2018). Images were first corrected for physiological motion using the Analysis of Functional Neuroimages (Cox, 1996). All subsequent preprocessing steps were conducted using the statistical parametric mapping software SPM12 (Penny et al., 2011). Preprocessing involved slice timing correction, spatial realignment, and coregistration with a resampled 3 mm isotropic voxel size, with no spatial smoothing. The mean time courses from participant-specific white matter and CSF masks were regressed out of the functional images, alongside estimates of the 6 rigid body motion parameters from each EPI run. To further correct for the effects of motion which may persist despite standard processing (Power et al., 2012), an additional motion scrubbing procedure was added to the end of our preprocessing pipeline (Campbell et al., 2013). Using a conservative multivariate technique, time points that were outliers in both the six rigid-body motion parameter estimates and BOLD signal were removed, and outlying BOLD signal was replaced by interpolating across neighboring data points. This method further reduces effects of motion-induced spikes on the BOLD signal without leaving sharp discontinuities because of the removal of outlier volumes.

Analysis. We used two main representational similarity analyses (Fig. 1D). To maximally benefit from the temporal resolution afforded by fMRI, paths were discretized into steps: each step corresponded to a TR, during which an entire brain volume was measured. In the first analysis, we computed the correlation between every given step (TR) and the average of all future steps (TRs) within a particular horizon (e.g., mean of future 10 TRs following the current TR). In the second analysis, following the equations for predictive or successor representations (Dayan, 1993; Momennejad et al., 2017), we computed the correlation between every given step and the discount-weighted sum of future steps within a horizon. The pattern across voxels at each future TR was weighed
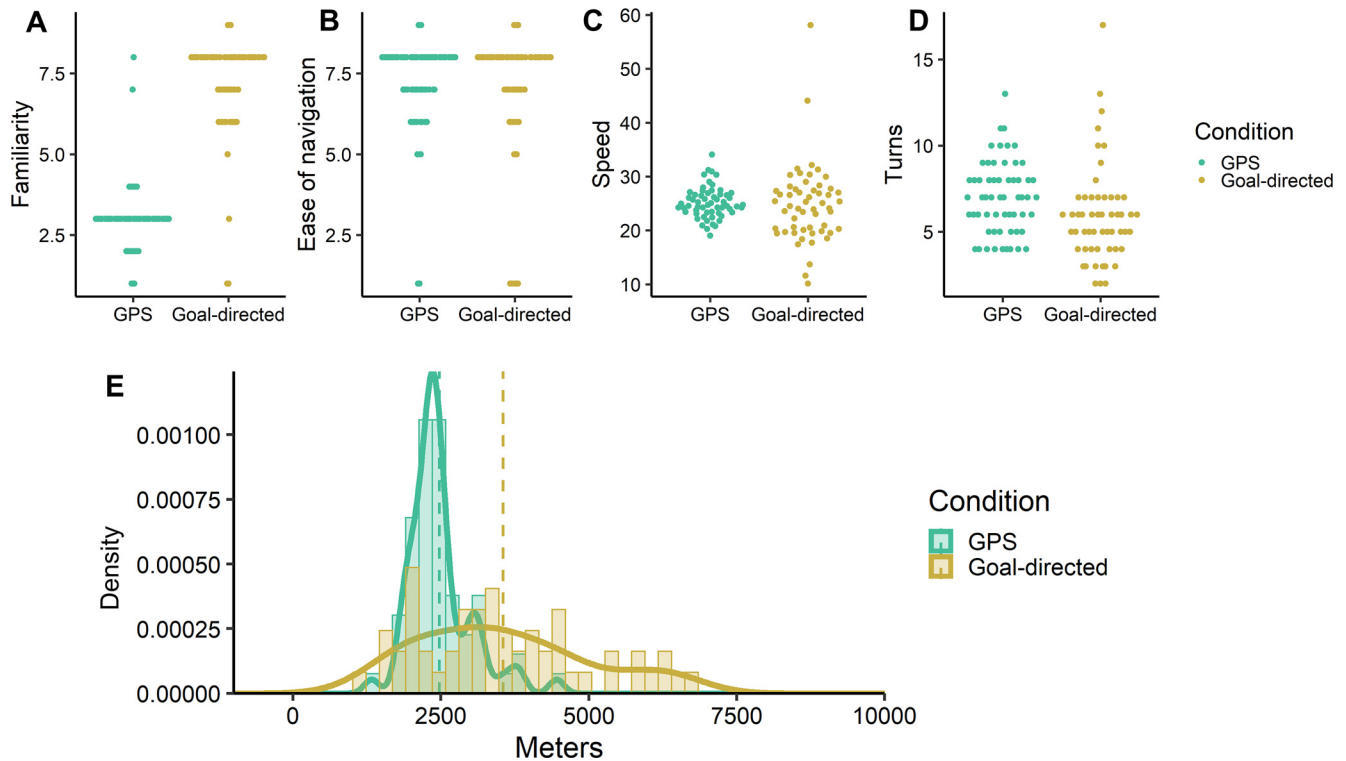
Figure 2. Descriptive statistics for navigated distances in goal-directed and GPS conditions. *A*, The goal-directed routes were rated as more familiar by participants than GPS routes. Goal-directed and GPS routes were matched in (*B*) ease of navigation and (*C*) speed of travel. *D*, GPS routes included more turns, on average, than goal-directed routes, but (*E*) the goal-directed routes tended to be longer than GPS routes.

exponentially using a discount parameter (i.e., γ value, γ) between 0 and 1, and the value of the discount parameter corresponded to the scale of abstraction, corresponding to different levels of a representational hierarchy (Momennejad and Howard, 2018).

ROI analysis. We investigated the predictive similarity of each state to future representations in a set of ROIs. To do so, we first extracted voxel-wise time courses across each navigated route and z-scored the values within each voxel. We then ran two predictive similarity analyses. First, we measured the correlation of each time point (TR) with the mean of successor TRs within a given horizon (e.g., correlation between TR at time t, and the mean of 10 following TRs). Second, we correlated the voxelwise pattern at each time point (TR) within each navigated route with a discount-weighted sum of future TRs. The patterns at future TRs were weighted by different constant values (γ), corresponding to different predictive spatial scales. The specified γ values were 0.1, 0.6, 0.8, and 0.9 (Fig. 1D). With increasing γ values, time points further in the future remain weighted >0.

As the average distance traversed within each TR was $\sim$25 m, a γ value of 0.1 meant that only each subsequent step (1 TR away) was weighted >0, and steps farther in the distance contributed little to no weight to the sum of future representations. We computed the predictive horizon using the unit of fMRI measurement (i.e., a TR of 2 s). Hence, depending on the speed of navigation, which was matched across conditions (see Fig. 2C), each step could cover a varying range of spatial distances (in meters) within and across subjects. Here we used the average distance traversed within a given horizon. For a γ value of 0.6, $\sim$7 steps in the future were weighted >0, corresponding to $\sim$175 m (see Fig. 6D). For a value of 0.8, $\sim$15 steps or 375 m were weighted >0, while this was the case for $\sim$32 steps or 800 m for a γ value of 0.9 (see Fig. 6D).

The TR-by-TR correlations within each route were averaged to derive the representation of future states on each trial. We first applied this analysis to *a priori* ROIs, including bilateral anterior and posterior hippocampi (aHPC, pHPC) and anterior and medial prefrontal cortical ROIs (antPFC, mPFC). We also examined the same measure in the mPFC and antPFC. The antPFC and mPFC ROIs were defined as

spheres surrounding peak voxels identified in preliminary findings from an fMRI adaptation of a known behavioral study of successor representations (Momennejad et al., 2017) reported by Russek et al. (2018). The spheres were centered on an anterior prefrontal voxel (MNI coordinates $x = 8$, $y = 68$, $z = 8$) and a medial prefrontal voxel (MNI coordinates $x = -22$, $y = 56$, $z = 10$). These analyses were performed for each of the ROIs, as well as a searchlight within the PFC.

PFC searchlight analysis. In order to identify any gradients of predictive representation in the PFC, a custom searchlight analysis was performed within a PFC mask (created in WFU PickAtlas). The analysis was restricted to gray matter voxels, and a spherical ROI with a 6 mm radius was used to iteratively correlate each TR with the discount-weighted sum of future states for voxels within each searchlight. The searchlight analysis was performed for four different values of γ: 0.1, 0.6, 0.8, and 0.9. The single-subject correlation maps were then compared against zero (AFNI *3dttest++*). The output z score maps were thresholded at values corresponding to 5% false positive rates established by a cluster-size permutation simulation (AFNI *ClustSim*).

Model-based analysis: the discount-weighted sum of successor states. This section addresses the reasoning behind testing the successor representation hypothesis in terms of pattern similarity between a given state and the discount-weighted sum of its successor states (Fig. 1). Consider an environment that consists of n states, some of which lead to one another. Consider T to be the $n \times n$ matrix of transition probabilities for one-step transitions among these n states. In a deterministic environment, when there is a transition from a given state S_i to state S_j, we assign 1 in the ith row and jth column of T. The successor representation under a random policy can be then computed from T as follows (for comparison to policy-dependent SR, see Momennejad, 2020):

$$M = (I - \gamma T)^{-1} \tag{1}$$

Equation 2 expands Equation 1 for computing the successor representation from state s_1 to the goal state s_g from T, which is one cell in the

SR matrix. Recall that T denotes the matrix of one-step transition probabilities among adjacent states, while SR contains multistep dependencies among nonadjacent states. Here the parameter t refers to the number of steps (or the distance) between states. This parameter need not denote temporal steps, and can denote any type of sequential relationship among states.

$$M(s_1, s_g) = \sum_{t=t1}^{tg} \gamma^t T(s_{t1}, s_t) \qquad (2)$$

Assume the starting state is s_1 and the goal state is s_5. Expanding Equation 2, the successor representation from States 1-5 is the fifth element in the first row of the successor representation (Eq. 3), and corresponds to the expected discounted number of times we expect to visit State 5, if we start from State 1:

$$M(s_1, s_5) = \gamma T(s_1, s_3) + \gamma^2 T(s_3, s_5) \qquad (3)$$

Equations 2 and 3 only capture 1 cell or element in the SR row associated with state s_1. In the successor representation framework, the *sth* row of the SR matrix (the M matrix in Dayan, 1993 equations) is the representation we expect to observe when the agent is in state s. It denotes how often we expect to visit the current state's successors on average and given a discount. A given row of the successor representation includes the present state, and the weighted representation of successor states. Thus, at the moment when an agent is in state s, the row activation of successor states predicts the simultaneous activation of γ-weighted representations. We take this simultaneous row activation as the sum of all activated weighted states in the row (Eq. 4).

$$Rep(s_i) = \sum_{i=t}^{n} \gamma^t Rep(s_t) \qquad (4)$$

In short, the first row of the SR matrix corresponds to the representation that is simultaneously activated when the agent is in State 1, which is the sum of $M(s_1, s_2)$, $M(s_1, s_3)$, $M(s_1, s_4)$, $M(s_1, s_5)$. Since we only have a goal-directed trajectory, this can be the weighted sum of representations of successor states (Eq. 4). Each successor state is weighted by the discount factor (γ, ɣ) to the power of its distance (here in the number of states) to the starting state. A simple prediction following this weighted sum view is that being in a given state along the trajectory activates the row associated with that state and hence the weighted sum of successor states on that trajectory. This predicts neural similarity between the current state and the weighted sum of successor state representations.

We did not have access to pretraining representations of the stimuli, for example, the uncorrelated representation of each location on the trajectory before being associated with specific paths (through lived experience in Toronto). Since we do not have these pretraining representations, this method offers an approximation of the expected similarity structure. Therefore, as a general rule, we make the following prediction. In a goal-directed trajectory, and assuming the agent stays on path, we can assume that the transition probability between two adjacent states, for example, $T(s_i, s_j)$, equals 1 (i.e., we have a deterministic Markov Decision Process (MDP)). We predict that Equation 3 approximates the pattern similarity of the TR in the *ith* state to the weighted sum of TRs that are its successor states. The predictive horizon is the successor distance within which the discount parameter $\gamma > 0$ (see Fig. 6). We hypothesize that different parts of the brain will show pattern similarity contingent with different values of the discount parameter $0 < \gamma < 1$, and thus different predictive horizons.

This is a first step toward testing the multiscale predictive representation hypothesis in a realistic navigation setting. To improve prediction accuracy, future studies are needed that incorporate diverse paths through each state, to each goal, and to different goals. These studies should include a larger graph or MDP of the environment with different starting and goal locations. In order to study map-dependent and path-dependent changes in the representation of each location, a study design is needed where the participants learn a new environment. Such studies would enable us to compare pretraining and post-training neural correlations among the states or locations in the environment.

Results

Participants navigated a set of distances they regularly traversed in everyday life (M-$_\text{goal-directed}$ = 3.5, M-$_\text{GPS}$ = 2.5 km). After completing each route, participants rated how familiar each route felt, and how difficult they found it to navigate on a scale from 1 to 9 (where 1 would correspond to least familiar and most difficult, respectively). As expected, the average reported familiarity was higher in the goal-directed condition (mean = 7.0, SD = 1.44) than in the GPS condition (mean = 3.0, SD = 0.51; $t_{(18)}$ = −10.53, $p < 0.001$, $d = 2.42$; Fig. 2A). The subjective difficulty was similar in the goal-directed (mean = 6.89, SD = 1.43) and GPS (mean = 7.24, SD = 1.08) conditions, suggesting that all navigated routes were perceived to be similarly undemanding ($t_{(18)}$ = 0.827, $p = 0.419$, $d = 0.190$; Fig. 2B). There was also no difference in movement speed across the goal-directed (mean = 24.91, SD = 7.66) and GPS conditions (mean = 25.23, SD = 2.25; $t_{(18)}$ = 0.191, $p = 0.851$, $d = 0.044$; Fig. 2C). GPS routes did, however, include more turns (mean = 7.08, SD = 1.39) than goal-directed routes (mean = 5.86, SD = 1.78; $t_{(18)}$ = 3.04, $p = 0.007$, $d = 0.698$; Fig. 2D). This was the case despite the GPS routes being shorter than goal-directed routes, on average ($t_{(18)}$ = −4.31, $p < 0.001$, $d = 0.989$; Fig. 2E).

Hippocampal and prefrontal gradients of near-future predictive representations

To investigate predictive representations along hippocampal and prefrontal hierarchies, we conducted a progression of analyses. First, we investigated representational similarity between each time point (TR) and the average of future n TRs, where n determined different future horizons (i.e., unweighted average of 1, 2, 3, 4, 5, or 10 future TRs) (Fig. 3). We conducted the analyses separately on 6 a priori ROIs of anterior-posterior hippocampus (split into 6 slices as in Brunec et al., 2018) and a priori selected mPFC and antPFC ROIs (see Materials and Methods). Second, we conducted the same analyses with discount-weighted sums of future TRs at different horizons (Figs. 4 and 5), focusing on two posterior and anterior hippocampal ROIs and mPFC and antMPFC. In follow-up analyses, we included the path distance on each route as a factor in the model. Third, we then conducted the discount-weighted sum RSA in a PFC-masked searchlight analysis to detect scales of representation in the PFC (Figs. 6 and 7).

Representational similarity to mean of future TRs across horizons

We conducted linear mixed effects models on these similarity measures in bilateral hippocampi for each of the routes traveled within each condition. We included average Fisher's z-transformed similarity on each route as the dependent variable, and axial segment (1-6), number of TRs (1-5), and hemisphere (L, R) as fixed effects. Similar analyses were performed for PFC ROIs. Participants were included as a random effect. The random intercept mixed effects models were implemented in R (R Core Team) using the packages *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) to assess significance. This produced a Type III ANOVA table with Satterthwaite's method of approximating degrees of freedom. Where these included decimal numbers, they were rounded to the nearest integer. Effect sizes for individual factors in mixed effects models were calculated as η_P^2 values using the *effectsize* R package (Ben-Shachar et al., 2020).

For overall model fits, we report marginal pseudo-R^2 values using the *r.squaredGLMM* function from the *MuMIn* R package, which represent the variance explained by fixed effects in the model (R^2_M) (Nakagawa et al., 2017; Bartoń, 2020). The similarity values for 10 TRs ahead were not entered in the present model because of the nonlinear shift from 5 to 10 TR, but they are plotted in Figure 3. All plots were generated with the ggplot2 package (Wickham, 2016).

Hippocampal results
We found a significant effect of axial segment ($F_{(5,6796)} = 45.38$, $p < 0.001$, $\eta^2_P = 0.03$), driven by greater future representations in the anterior segments compared with posterior ones. There was also a main effect of condition ($F_{(1,6796)} = 1182.35$, $p < 0.001$, $\eta^2_P = 0.15$), reflecting generally greater values in the goal-directed (Fig. 3A), compared with the GPS condition (Fig. 3B), and a significant effect of the future horizon ($F_{(1,6796)} = 633.44$, $p < 0.001$, $\eta^2_P = 0.27$), reflecting higher similarity values for states closer to the present. There was a main effect of hemisphere, reflecting higher values in the right compared with the left hemisphere ($F_{(1,6796)} = 6.97$, $p = 0.008$, $\eta^2_P = 0.001$). There were significant interactions between axial segment and condition ($F_{(5,6796)} = 6.97$, $p < 0.001$, $\eta^2_P = 0.004$), axial segment and future horizon ($F_{(20,6796)} = 2.13$, $p = 0.002$, $\eta^2_P = 0.006$), and condition and future horizon ($F_{(4,6796)} = 13.16$, $p < 0.001$, $\eta^2_P = 0.008$). The latter interaction is of particular interest as it suggests that the decline across different predictive horizons was greater in the GPS compared with the goal-directed condition. There was no significant three-way interaction ($F < 1$). The overall R^2_M of the model was 0.30.

PFC results
We fit the same models separately for the *a priori* selected ROIs in antPFC and mPFC. In the antPFC (overall $R^2_M = 0.22$), there was a significant main effect of condition ($F_{(1,1222)} = 363.76$, $p < 0.001$, $\eta^2_P = 0.23$), as well as a main effect of future horizon ($F_{(4,1222)} = 48.36$, $p < 0.001$, $\eta^2_P = 0.14$), but no condition by future horizon interaction ($F_{(4,1222)} = 1.18$, $p = 0.319$, $\eta^2_P = 0.004$). In the mPFC (overall $R^2_M = 0.26$), there was a significant effect of condition ($F_{(1,1222)} = 218.77$, $p < 0.001$, $\eta^2_P = 0.19$) and a significant effect of future horizon ($F_{(4,1222)} = 114.82$, $p < 0.001$, $\eta^2_P = 0.27$), but again no significant condition by future horizon interaction ($F_{(4,1222)} = 1.82$, $p = 0.122$, $\eta^2_P = 0.006$).

Comparing the representational similarity in the goal-directed and GPS conditions against zero, we found that the antPFC displayed >0 similarity for every predictive horizon, including 10 steps ahead, in the goal-directed condition (all *p* values < 0.001), but only up to 5 steps in the GPS condition (all *p* values for 1-5 steps < 0.001). In contrast, the mPFC only displayed >0 similarity up to 5 steps in the future on goal-directed routes (*p* values < 0.001) and 3 steps on GPS routes (*p* values ≤ 0.002). The
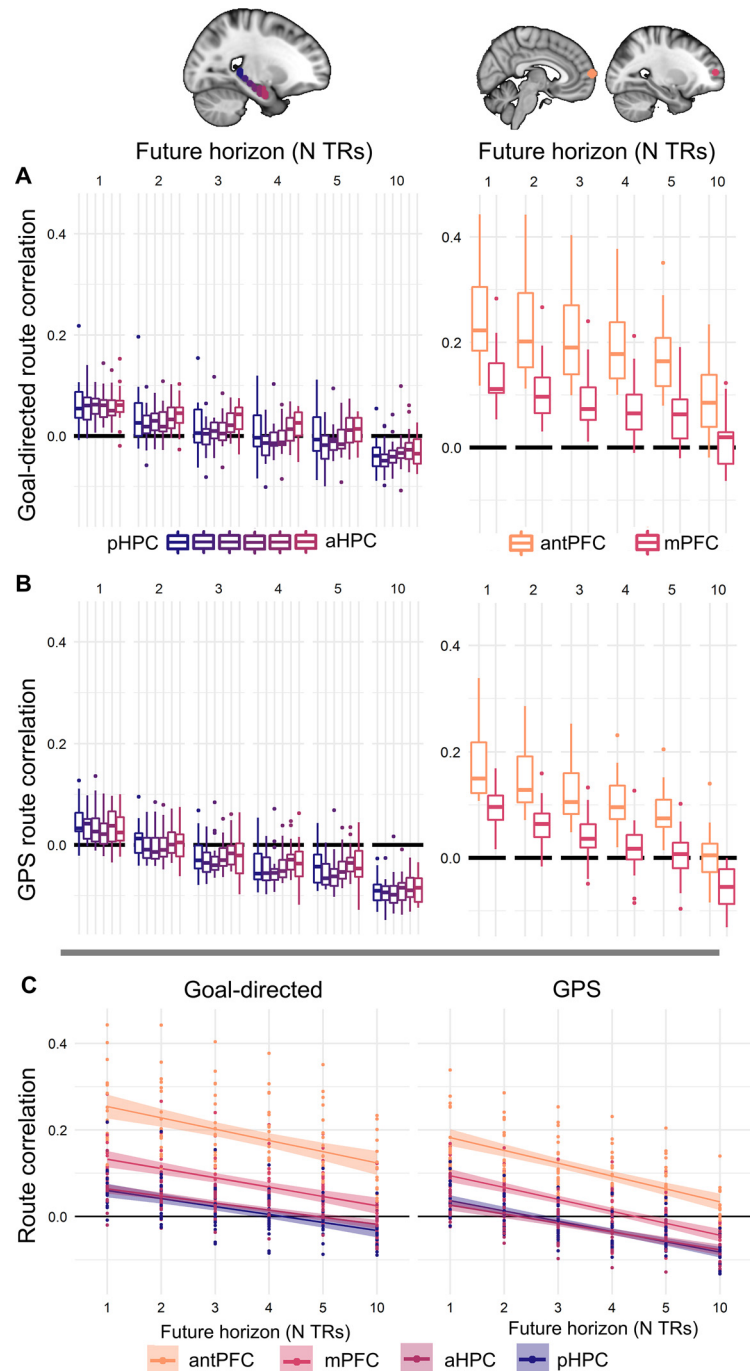

Figure 3. Similarity of each TR to mean of future TRs (equally weighted). Average correlation between each time point and the average of future 1/2/3/4/5 or 10 time points in the (*A*) the goal-directed condition and (*B*) the GPS condition. *C*, Representational similarity for all ROIs across all temporal lags. More posterior regions cross zero at smaller horizons. The difference between the anterior and posterior hippocampus is less pronounced here than in subsequent analyses where neural representations were weighted by different discount factors. The aHPC and pHPC labels refer to anterior-most and posterior-most hippocampal segments, respectively, shown in *A* and *B*.

anterior-most hippocampal segment displayed >0 similarity for up to 4 steps in the future (*p* values ≤ 0.006) on goal-directed routes and only 1 step on GPS routes (*p* < 0.001), while the posterior-most hippocampal segment displayed >0 similarity for 1 step on goal-directed routes (*p* < 0.001), and 2 steps on GPS routes (*p* values ≤ 0.006).

We next conducted similar analyses with the weighted sum of future TRs of different horizons.
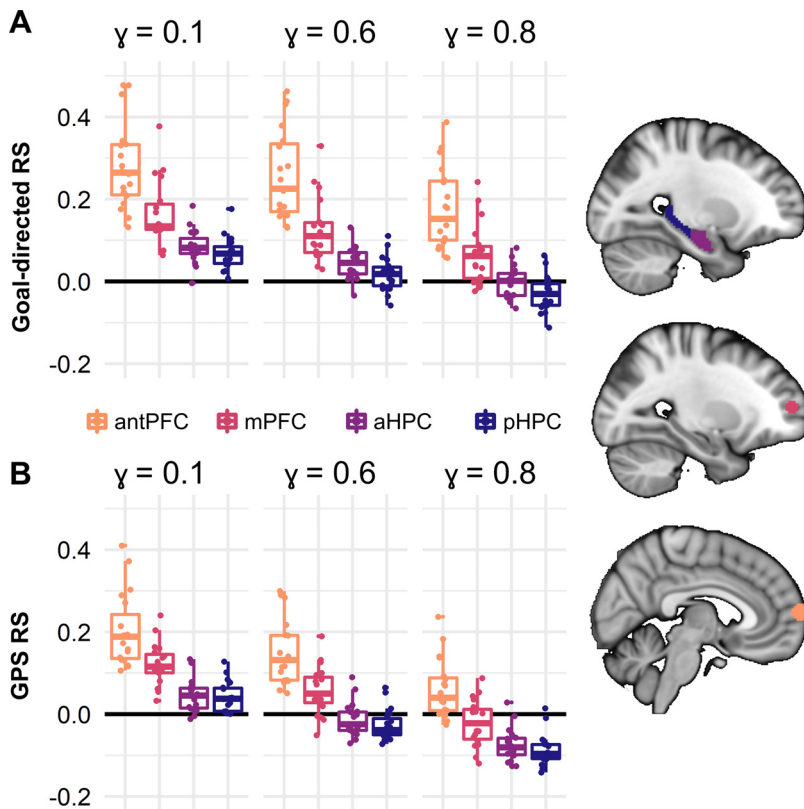
A

B

Figure 4. Predictive similarity across predictive scales. Correlations between current time points and the γ-weighted sum of future states for different values of γ, in the four specified ROIs in the (*A*) goal-directed and (*B*) GPS conditions. γ = 0.1 only included 1 step (1 TR) away, γ = 0.6 reached ~6 or 7 steps in the future, corresponding to ~175 m, γ = 0.8, ~14 steps or 350 m ahead.

Model-based representational similarity to future TRs in ROIs
When an RL agent is in a given state during navigation, the discount-weighted sum of the successor states is the predictive representation for that state (see Materials and Methods). Therefore, we investigated the similarity between each time point and γ-weighted sum of the representations of future states. We ran a series of linear mixed effects models following the logic described above, including each route within each of the conditions. The models included Fisher's z-transformed representational similarity values as the dependent variable, with γ and condition as fixed effects and participant as a random effect. γ was modeled as an ordinal variable. All analyses were implemented using the same packages as above. For the hippocampus, the reported statistics and plotted values apply to the right hippocampus, as there was no significant difference between left and right hippocampi (all p values > 0.34).

Mixed effects analysis
The first mixed effects model included all ROIs to compare average representational similarity differences across regions with different hypothesized scales. There was a significant main effect of γ ($F_{(2,1448)} = 322.14$, $p < 0.001$, $\eta^2_P = 0.31$), suggesting (not surprisingly) more representational similarity within horizons that are closer to the present state. We also observed a significant main effect of condition ($F_{(1,1452)} = 309.46$, $p < 0.001$, $\eta^2_P = 0.18$), suggesting representational similarity at higher predictive horizons in the goal-directed compared with the GPS condition (Fig. 4*A,B*). There was a main effect of ROI ($F_{(3,1448)} = 547.38$, $p < 0.001$, $\eta^2_P = 0.53$), confirming the hypothesis of longer predictive horizons in the

antPFC, followed by mPFC, aHPC, and pHPC. There was also a significant interaction between γ and condition ($F_{(2,1448)} = 7.49$, $p < 0.001$, $\eta^2_P = 0.01$), and a significant interaction between condition and ROI ($F_{(3,1448)} = 10.13$, $p < 0.001$, $\eta^2_P = 0.02$). The overall R^2_M of the model was 0.57.

Within-ROI analyses
Follow-up mixed effects models were conducted for predictive similarity values within each ROI. Significance was established against a Bonferroni-adjusted value of α = 0.0125 (for 4 ROIs). In the antPFC (overall R^2_M = 0.27), there was a significant main effect of γ ($F_{(2,347)} = 53.29$, $p < 0.001$, $\eta^2_P = 0.23$). There was also a significant effect of condition, with significantly higher correlations in the goal-directed than the GPS condition ($F_{(1,349)} = 103.42$, $p < 0.001$, $\eta^2_P = 0.23$). There was no significant γ × condition interaction ($F < 1$). In mPFC (overall R^2_M = 0.34), there was again a significant main effect of γ ($F_{(2,350)} = 106.39$, $\eta^2_P = 0.38$), as well as a main effect of condition ($F_{(1,352)} = 83.19$, $p < 0.001$, $\eta^2_P = 0.19$) in the same direction as the antPFC. There was no significant γ × condition interaction ($F_{(2,350)} = 3.44$, $p = 0.033$, $\eta^2_P = 0.02$).

In the aHPC (overall R^2_M = 0.45), there was a significant main effect of γ ($F_{(2,348)} = 151.90$, $p < 0.001$, $\eta^2_P = 0.47$), a main effect of condition ($F_{(1,350)} = 128.05$, $p < 0.001$, $\eta^2_P = 0.27$), as well as a γ × condition interaction ($F_{(2,348)} = 4.89$, $p = 0.008$, $\eta^2_P = 0.03$). As in the mPFC, this interaction reflected a steeper slope across γ values in the GPS condition (–0.16) than in the goal-directed condition (–0.12). In the pHPC (overall R^2_M = 0.47), there was a significant main effect of γ ($F_{(2,349)} = 218.38$, $p < 0.001$, $\eta^2_P = 0.56$), a main effect of condition ($F_{(1,351)} = 87.99$, $p < 0.001$, $\eta^2_P = 0.20$), and a significant γ × condition interaction ($F_{(2,349)} = 3.81$, $p = 0.023$, $\eta^2_P = 0.02$), again reflecting a steeper slope in the GPS condition (–0.17), compared with the goal-directed condition (–0.13).

To test for evidence of predictive representations, we conducted one-sample t tests to test these values against zero, with an adjusted value of α = 0.002 (24 comparisons in total). At γ = 0.1, the similarity values in all ROIs were significantly >0 in both conditions. At γ = 0.6, similarity values for all ROIs but the pHPC were significantly >0 in the goal-directed condition. In the GPS condition, however, similarity values in neither the aHPC nor the pHPC were significantly >0. At γ = 0.8, values in both antPFC and mPFC remained significantly >0 in the goal-directed condition, but only antPFC remained >0 in the GPS condition. For this value of γ, the values in aHPC and pHPC were not significantly >0 in either condition, and were actually significantly <0 in the pHPC. This significant negative correlation could reflect the differentiation of neural patterns across time, potentially as a manner of separating experience into fine-grained units.

Representational similarity during goal-directed navigation is related to traveled path distance
If the hippocampus and PFC represent planning processes associated with the currently navigated route, these representations should be modulated by the route path distance. To test this, we included the path distance on each route as a factor in the mixed

effects model. Path distance was calculated as the summed change in longitude and latitude coordinates between each adjacent pair of TRs. To account for the contribution of time, we also regressed out the number of TRs on each route. The reported model-fits thus account for the variability in the amount of time spent navigating different routes. Before running these models, we mean-centered distances within each participant to account for different ranges traveled. We excluded nine goal-directed routes from a total of 8 participants because of improbably long paths that diverged >1.5 km from the main path. Including these paths, however, did not change the significance of any of the results.

Path distance results
In the goal-directed condition (overall $R^2_M = 0.71$), there were significant effects of ɣ ($F_{(2,630)} = 186.83$, $p < 0.001$, $\eta^2_p = 0.37$) and ROI ($F_{(3,630)} = 369.49$, $p < 0.001$, $\eta^2_p = 0.64$). There was no significant main effect of path distance ($F_{(1,645)} = 1.06$, $p = 0.304$, $\eta^2_p = 0.002$), but there were significant interactions between ROI and path distance ($F_{(3,630)} = 38.13$, $p < 0.001$, $\eta^2_p = 0.15$) and ɣ and path distance ($F_{(2,630)} = 6.47$, $p = 0.002$, $\eta^2_p = 0.02$; Fig. 5). The plotted values in Figure 5 were estimated using the effects package in R (Fox, 2003; Fox and Weisberg, 2011). There was no significant interaction between ɣ and ROI, nor a three-way interaction (both p values > 0.30). As predicted, we observed no main effect of path distance in the GPS condition ($F < 1$), nor any interactions with ROI ($F_{(3,671)} = 2.07$, $p = 0.103$, $\eta^2_p = 0.009$) or ɣ ($F < 1$). The main effects of ɣ ($F_{(2,671)} = 180.62$, $p < 0.001$, $\eta^2_p = 0.35$) and ROI ($F_{(3,671)} = 209.62$, $p < 0.001$, $\eta^2_p = 0.48$) remained significant, however. The overall R^2_M of this model was 0.68.

ROI and path distance interactions
We conducted a linear mixed effects model for each of the ROIs, predicting representational similarity from path distance and ɣ. In the antPFC (overall $R^2_M = 0.32$), there were significant effects of ɣ ($F_{(2,144)} = 33.92$, $p < 0.001$, $\eta^2_p = 0.32$) and path distance ($F_{(1,149)} = 91.51$, $p < 0.001$, $\eta^2_p = 0.38$), but no interaction between the two ($F_{(2,144)} = 1.09$, $p = 0.340$, $\eta^2_p = 0.01$). This suggests that the effect of path distance was stable across different predictive horizons in antPFC. In mPFC (overall $R^2_M = 0.32$), the effects of ɣ ($F_{(2,144)} = 71.77$, $p < 0.001$, $\eta^2_p = 0.50$) and path distance ($F_{(1,147)} = 82.70$, $p < 0.001$, $\eta^2_p = 0.36$) were again significant, as was the interaction between the two ($F_{(2,144)} = 3.62$, $p = 0.029$, $\eta^2_p = 0.05$). In the aHPC (overall $R^2_M = 0.40$), there was a significant effect of ɣ ($F_{(2,145)} = 69.82$, $p < 0.001$, $\eta^2_p = 0.49$), a significant effect of path distance ($F_{(1,152)} = 47.17$, $p < 0.001$, $\eta^2_p = 0.24$), and a weaker interaction between ɣ and path distance ($F_{(2,145)} = 2.93$, $p = 0.057$, $\eta^2_p = 0.04$). Finally, in the pHPC (overall $R^2_M = 0.50$), there were significant effects
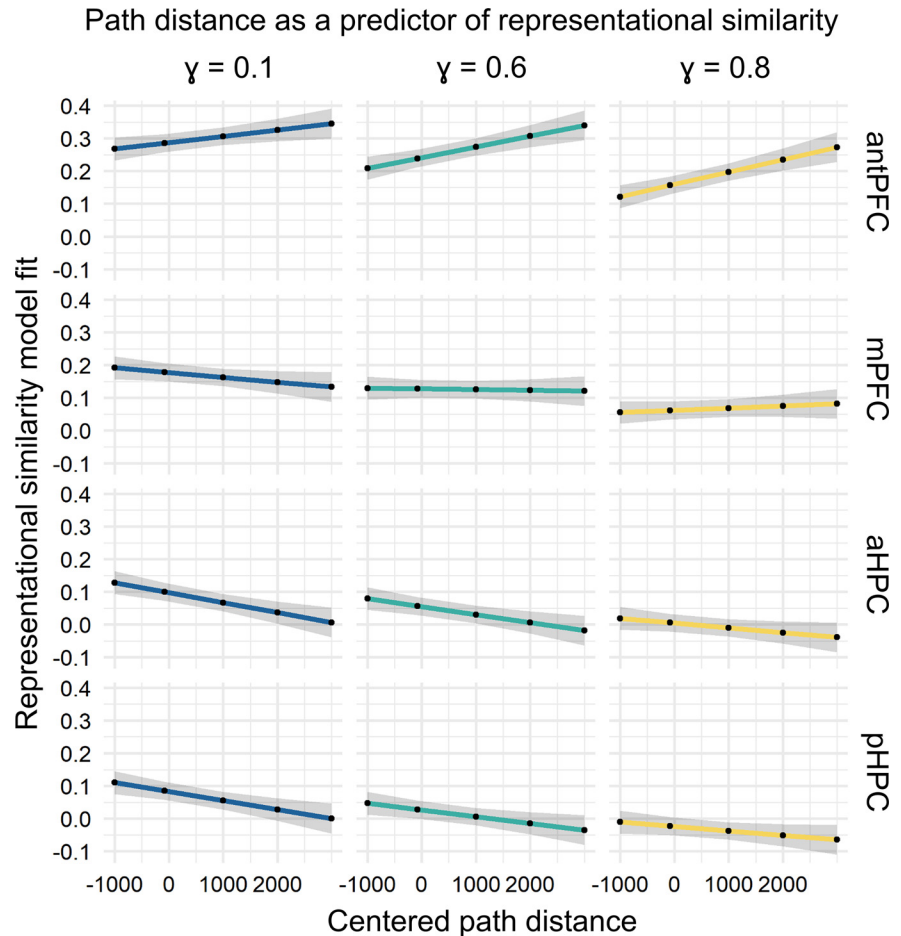
Path distance as a predictor of representational similarity

Figure 5. Linear mixed effects model predicting representational similarity (*y* axis) from path distance (*x* axis), ɣ, and ROI. Voxelwise patterns in different ROIs interacted differently with path distance: in the antPFC, routes with longer path distances were associated with greater representational similarity, whereas the opposite trend was present in the hippocampus (both aHPC and pHPC). Plot represents the model fit values and CIs. These reflect the relationships between the variables of interest after regressing out the effect of the number of TRs on each route and accounting for all other main effects and interactions.

of ɣ ($F_{(2,144)} = 146.40$, $p < 0.001$, $\eta^2_p = 0.67$), path distance ($F_{(1,148)} = 82.52$, $p < 0.001$, $\eta^2_p = 0.36$), and a weaker interaction between the two ($F_{(2,144)} = 3.23$, $p = 0.042$, $\eta^2_p = 0.04$).

Comparison with Euclidean distance
To establish how specific these results were to the traversed paths, we reran the models but this time included the Euclidean distance from start to goal as a predictor instead. In the goal-directed condition (overall $R^2_M = 0.69$), the effects of ɣ and ROI remained significant (both p values < 0.001), but there was no main effect of Euclidean distance ($F < 1$), and no significant interaction between ɣ and Euclidean distance ($F_{(2,631)} = 1.93$, $p = 0.145$, $\eta^2_p = 0.006$). There was an interaction between ROI and Euclidean distance ($F_{(3,630)} = 3.33$, $p = 0.019$, $\eta^2_p = 0.02$), but no three-way interaction ($F < 1$). In the GPS condition, the effects of ɣ and ROI were again significant (p values < 0.001), and there was a weaker main effect of Euclidean distance ($F_{(1,49)} = 4.39$, $p = 0.041$, $\eta^2_p = 0.08$), but no other main effects or interactions (all p values < 0.60).

Model-based representational similarity in prefrontal searchlights
PFC has a much larger volume than the hippocampus. In order to identify hierarchies of predictive representations
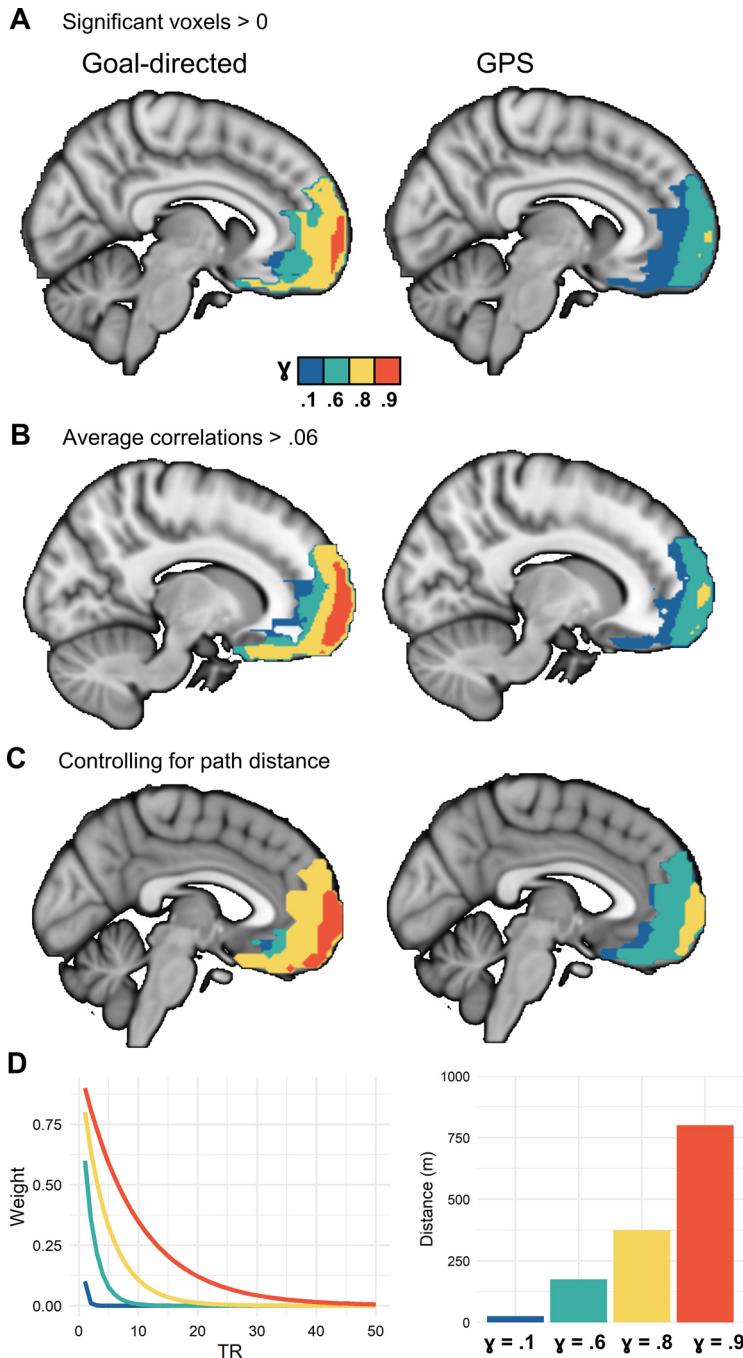
A Significant voxels > 0

B Average correlations > .06

C Controlling for path distance

D

Figure 6. One-sample t tests for goal-directed and GPS condition. **A**, Voxels with significant representation of future states in the goal-directed and GPS conditions using a one-sample t test against zero. **B**, Voxels with representational similarity (correlation) values >0.06 for each value of ɣ. **C**, One-sample t tests with distance as covariate. The results look very similar to running a t test on goal-directed routes versus zero, and GPS routes versus zero. The mean distance, per participant, per condition, was included as a covariate. **D**, Discounted weights corresponding to different gammas were applied to each successor TR. The average distance covered in each TR was ~25 m (24.8 m). Based on this, we computed approximate distances corresponding to predictive horizons for each discount parameter. The exact distances for each discount parameter differed across routes and participants depending on their speed. ɣ = 0.1 only included 1 step (1 TR) away, ɣ = 0.6 reached ~7 steps in the future, corresponding to ~175 m, ɣ = 0.8, ~15 steps or 375 m, ɣ = 0.9 reached ~32 steps or 800 m ahead.

comparable to hippocampal ROIs, we ran a searchlight analysis and computed similarity for voxels within every spherical searchlight (of 6 mm radius). The searchlight analysis was performed for four values of ɣ (0.1, 0.6, 0.8, 0.9) within each of the conditions. The thresholded z score maps for different values of ɣ are displayed as overlays in

Figure 6A, along with the average thresholded similarity maps within each condition (thresholded at 0.06; Fig. 6B).

Prefrontal hierarchy
To capture the gradient of values from the anterior-most to the posterior-most segments of the PFC, we calculated the average value of representational similarity across voxels within each anterior-posterior slice (i.e., the y direction). The slopes are plotted in Figure 7. These plots reveal a gradation of predictive representations extending from posterior-most to anterior-most slices of the PFC. This trend was reliable in both the goal-directed and GPS conditions, but the representational similarity values were consistently greater in the goal-directed condition.

To account for the proportion of different histologically defined brain regions covered by each significant cluster, we calculated the percentage of overlap between each prefrontal Brodmann area (BA) region and the significant voxels for each value of ɣ in each of the conditions (Table 1; Fig. 8). These percentages represent the proportion of each BA region covered by the significant thresholded clusters. We found the largest overlap between voxels in the antPFC (BA 10) and significant voxels in the searchlight analysis with various ɣ values. Following anterior and polar PFC was BA 11, corresponding to the OFC, and then BA 25 and 32, corresponding to subgenual area or cingulate cortex and ACC, respectively. These regions were followed by smaller overlap in area 47, corresponding to the orbital part of the inferior frontal gyrus, areas 46 and 9 corresponding to the dorsolateral PFC, and no overlap in area 45 corresponding to the inferior frontal gyrus.

Representational similarity slope along PFC hierarchy
Controlling for distance: matched distance analysis
As discussed in earlier sections, the distances were not matched between the two conditions (Fig. 2E). To account for this difference, we conducted a matched analysis in which we manually selected pairs of routes with the minimum difference in distance for each participant, up to 1 km (Fig. 9A). We were unable to include 3 of the participants in this analysis as the distances in their goal-directed and GPS routes were too different (with a difference in distance > 1.5 km). For the remaining 16 participants, there was no significant difference between the selected GPS and goal-directed routes ($p = 0.215$). We ran a paired-samples t test comparing participants' prefrontal RSA maps for the two selected routes. We also included the difference in distance for the two selected routes as a covariate
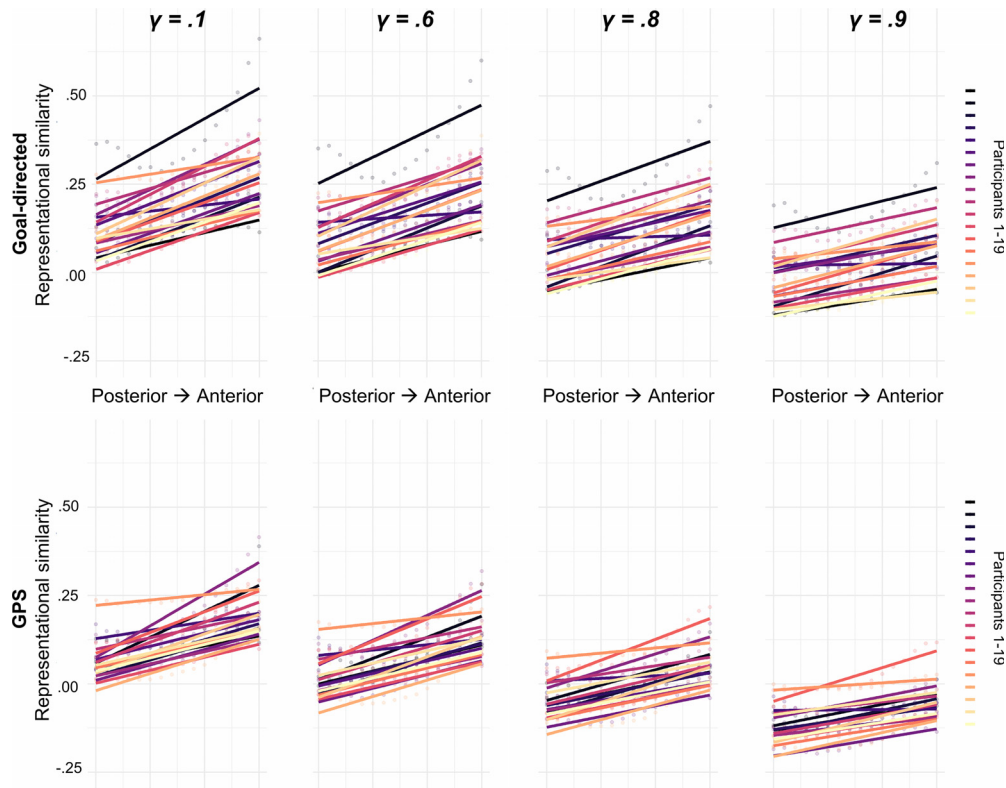
Figure 7. Increasing predictive similarity along posterior PFC to antPFC. In order to indicate which PFC regions displayed higher predictive similarity, we computed the slope of correlations for posterior PFC to antPFC slices for goal-directed and GPS conditions. We computed these slopes for 4 values of ɣ, corresponding to gradients of low to high scales. Each line indicates predictive similarity results from 1 of 19 participants.

for each participant. The brain maps of the average correlation values thresholded at 0.04 are presented in Figure 9B, and the results of the 5% False Positivity Rate (FPR)-corrected t test in Figure 9C.

We compared matched-distance searchlight results in the goal-directed and GPS conditions. In this comparison, relatively few clusters significantly differed between the goal-directed and GPS conditions. However, the comparison at each level of ɣ suggests that there is a set of clusters along the rostrocaudal extent of the PFC which differentiates between goal-directed and GPS-guided navigation (Table 1). Notably, while only orbitofrontal clusters were significantly different for smaller horizons, more dorsal and rostral/polar PFC clusters emerged in the comparison of larger horizons or scales, between the goal-directed and GPS conditions. It is worth noting, however, to ensure matched distances between the goal-directed and GPS condition, we excluded individuals with a large difference between the distances in the two conditions. As a result, this analysis only included individual paths from 16 participants, which likely results in increased noise and lower statistical power, which is common when using more naturalistic data.

Discussion

We investigated the hypothesis that relational knowledge, about navigational paths, is organized as multiscale predictive representations in hippocampal and prefrontal hierarchies. We found evidence for such multiscale representations in a task where participants navigated the city of Toronto virtually in goal-directed and GPS-guided conditions, with realistically long distances. Our fMRI representational similarity analysis between each state

(TR) and a discounted sum of its prospective states (at multiple scales, 25-875 m) confirmed this hypothesis. These results support the idea that prefrontal-hippocampal representations organize relational knowledge, in this case for navigation, at different scales of generalization and abstraction (Behrens et al., 2018; Momennejad and Howard, 2018).

Our primary goal was to investigate at which scale, and in which condition, different brain regions remain informative about the upcoming path. Representational similarity patterns in different regions stopped carrying information about predicted paths at different horizons. Namely, for higher horizons, fewer regions had above chance similarity to planned paths (Fig. 3), suggesting a hierarchy of representation. Moreover, representations in posterior regions stopped being predictive at smaller scales, and gradually more anterior regions remained predictive at longer horizons (Fig. 4). Notably, there was an interaction between hierarchy and condition: in the goal-directed condition (Fig. 4A), the findings were more pronounced than in the GPS-guided condition (Fig. 4B). This finding reveals that during planning at realistically long horizons regions higher in the representational hierarchy carry predictive information.

We have reported four main findings. First, fMRI similarity reflected longer predictive horizons for paths in the goal-directed, compared with the GPS condition. Second, similarity in the anterior hippocampus and antPFC was significantly higher in the goal-directed condition and for longer horizons (Fig. 4). Third, predictive representations were organized along a posterior-anterior hippocampal gradient of predictive horizons (25-175 m) with larger scales in gradually more anterior regions (Fig. 3). Fourth, representational similarity to future horizons was

Table 1. Proportion of each prefrontal BA accounted for by the significant prefrontal voxels[a]

	Goal-directed				GPS			
	ɣ = 0.1	ɣ = 0.6	ɣ = 0.8	ɣ = 0.9	ɣ = 0.1	ɣ = 0.6	ɣ = 0.8	ɣ = 0.9
BA9	5.8%	5.8%	5.6%	0%	5.8%	4.3%	0%	0%
BA10	59.4%	59.3%	54.9%	13.9%	59.1%	48.0%	4.9%	0%
BA11	46.5%	44.4%	32.6%	6.0%	42.6%	21.8%	0%	0%
BA25	41.1%	37.4%	15.9%	0%	21.5%	0%	0%	0%
BA32	23.9%	21.6%	11.2%	0%	21.1%	2.8%	0%	0%
BA47	16.6%	15.6%	5.8%	0%	12.7%	1.1%	0%	0%
BA46	6.8%	6.8%	6.8%	0%	6.8%	4.2%	0%	0%
BA45	0%	0%	0%	0%	0%	0%	0%	0%

[a] Results were driven from the one-sample *t* test results displayed in Figure 5A (not matched for distance). Proportions are displayed for each value of ɣ within each condition.
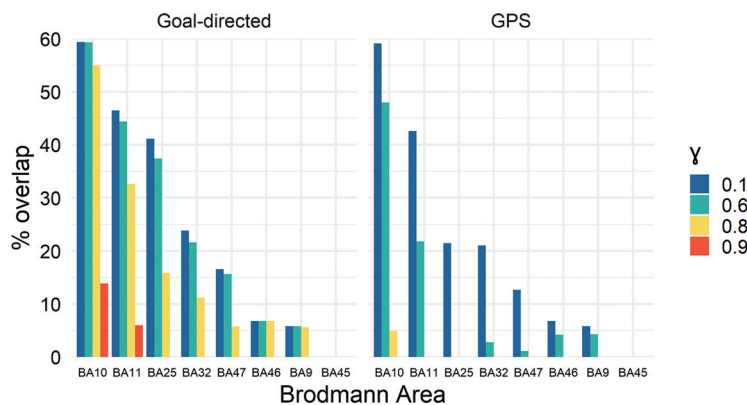

Figure 8. PFC hierarchy in the goal-directed and GPS conditions. Proportion of prefrontal BAs accounted for by the significant PFC voxels in searchlight analysis are shown. Results were driven from the one-sample *t* test results displayed in Figure 5A (not matched for distance). Color bars represent different discount values (ɣ) corresponding to different predictive horizons within each condition (blue: ɣ = 0.1; green: ɣ = 0.6; yellow: ɣ = 0.8; red: ɣ = 0.9).

organized along an anterior-posterior gradient in the PFC with larger-scale horizons (25-875 m) in gradually more anterior regions (Fig. 6).

Representational hierarchy
In spatial navigation, hierarchical representation could enable hierarchical planning and subgoal computation (Ribas-Fernandes et al., 2019). Our proposal is that larger and more abstract scales of predictive representations in antPFC may support planning at larger scales (Fig. 1, large-scale graph). This higher-level plan may be translated into gradually more granular representations in prepolar PFC, OFC, and anterior hippocampal regions (Fig. 1, mid-scale graph), and finer-scale trajectories are translated by hippocampal gradients to the smallest predictive horizons of place fields (Fig. 1, small-scale graph). In our analyses, this was reflected in a gradation of representational similarity: at longer horizons, fewer regions had above chance similarity to planned paths.

Hippocampal hierarchy
One possibility is that the PFC represents the global structure of each route, while the hippocampus supports fine-grained representation of individual locations. This is consistent with recent cognitive map work in rodents, monkeys, and humans indicating PFC's involvement in active navigation and planning (Epstein et al., 2017), as well as evidence that the dorsal-ventral or posterior-anterior hippocampal axis supports gradually larger spatiotemporal scales (Poppenk et al., 2013; Strange et al., 2014; Peer et al., 2019) and inference on mnemonic relations (Collin et al., 2015; Schlichting and Preston, 2015). Recent computational perspectives

suggest that the hippocampus and PFC form and update a predictive map of the state space at multiple scales (Momennejad and Howard, 2018), organizing relational knowledge of spatial and nonspatial states (McKenzie et al., 2014; Schuck et al., 2016; Garvert et al., 2017; Bellmund et al., 2018).

Prefrontal hierarchy
Comparing predictive similarity across the PFC (Figs. 6-9; Table 1) revealed an overall effect of condition and prefrontal gradient. Longer predictive similarity horizons were observed in the goal-directed versus GPS-guided condition, and antPFC regions showed predictive similarity for longer horizons: the anterior or polar PFC (BA 10; Table 1; Fig. 8), OFC (BA 11), and granular and ACC (BA 25 and 32), consistent with the slope of predictive similarity in Figure 7. Such a prefrontal hierarchy of relational abstraction could support task sets and schema. BA 10 is structurally well connected within the PFC and with the rest of the cortex and has long decay, thus a candidate for supporting higher scales of abstraction, from predictive representations with larger scales of integration to clustering relational graphs with higher radii. These functions may rely on longer sustained memory leading to binding over longer time scales, associating farther apart locations, or increasing representational similarity among clusters of associations.

Nonspatial relevance
Previous work has proposed hierarchies of predictive representations along prefrontal and hippocampal gradients (Momennejad and Howard, 2018), a hierarchy of time scales in the brain (Chen et al., 2015), and a role for hippocampal-prefrontal interactions in integrating episodes to build abstract schema (Schlichting and Preston, 2017). Similar representational hierarchy may also underlie relational knowledge and category generalization (Constantinescu et al., 2016), abstraction and transfer (Cole et al., 2011), reward predictions (Takahashi et al., 2017), associative inference, and schema learning (Moscovitch and Melo, 1997; Zeithamova and Preston, 2010; Zeithamova et al., 2012; van Kesteren et al., 2013; Hebscher and Gilboa, 2016; Spalding et al., 2018; Yu, 2018; Lee et al., 2021). A crucially nonspatial body of evidence indicates a functional role for antPFC in the encoding and retrieval of prospective memory task-sets and goals (Haynes and Rees, 2006; Gilbert, 2011; Momennejad and Haynes, 2012, 2013). Lesions to the frontopolar cortex do

not impair navigation, intelligence, or working memory, but impair multitasking and prospective memory (Burgess, 2000; Volle et al., 2011), such as completing a sequential plan for simple everyday tasks (Burgess, 2000). This fits with the proposal that the PFC is organized in a rostrocaudal hierarchy (Koechlin et al., 2003; Badre and D'Esposito, 2007; Koechlin and Hyafil, 2007; Koechlin, 2011), with more anterior or rostral regions corresponding to higher levels of integration and relational abstraction (Bunge et al., 2003; Christoff et al., 2009; Momennejad and Haynes, 2013).

OFC

More anterior OFC yielded higher predictive similarity for larger predictive horizons (Figs. 4, 8). OFC and antPFC have both been suggested to support model-based RL (Daw et al., 2011), where an agent unfolds a learned state-action-state associative model during goal-directed planning and decision-making (McDannald et al., 2012, 2014; Daw and Dayan, 2014; Pauli et al., 2019). Thus, OFC may maintain task-relevant state-state relational maps that enable iterative value computation in planning and decision-making (Daw et al., 2005; Simon and Daw, 2011; Keiflin et al., 2013; Wilson et al., 2014; Wimmer and Büchel, 2019). Predictive representations in anterior hippocampus were the most similar to OFC representations, consistent with recent work on OFC-hippocampal interactions in model-based behavior (Wood and Grafman, 2003; Keiflin et al., 2013; Schuck et al., 2016; Wikenheiser and Schoenbaum, 2016; Miller et al., 2017; Vikbladh et al., 2019).

Prefrontal hierarchies of representations need not be static. These representations could be constructed from compressed representation, for example, eigenvectors (Stachenfeld et al., 2017), inverse Laplace transform (Momennejad and Howard, 2018), or generative models (Whittington et al., 2020). Future studies can shed light on prefrontal and medial temporal contributions to information integration, eigen-decomposition, generative models, and abstraction.

Caveats and future directions

The fMRI dataset used here (Brunec et al., 2018) was acquired for different questions, leaving some caveats for present purposes, some of which we addressed in our control analyses, whereas others remain to be addressed by future studies.

The first caveat: navigated routes in the goal-directed condition were longer than those in the GPS-guided condition (Fig. 2). To overcome this, we controlled for distance in one-sample t tests to reveal regions with significant pattern similarity for a given horizon (Fig. 6C). In a more conservative analysis, we excluded longer routes from analyses, including only goal-directed routes within the range of distances in the GPS-guided condition (Fig. 9). Consistent with our earlier findings, longer predictive scales engaged more dorsal PFC regions in the goal-
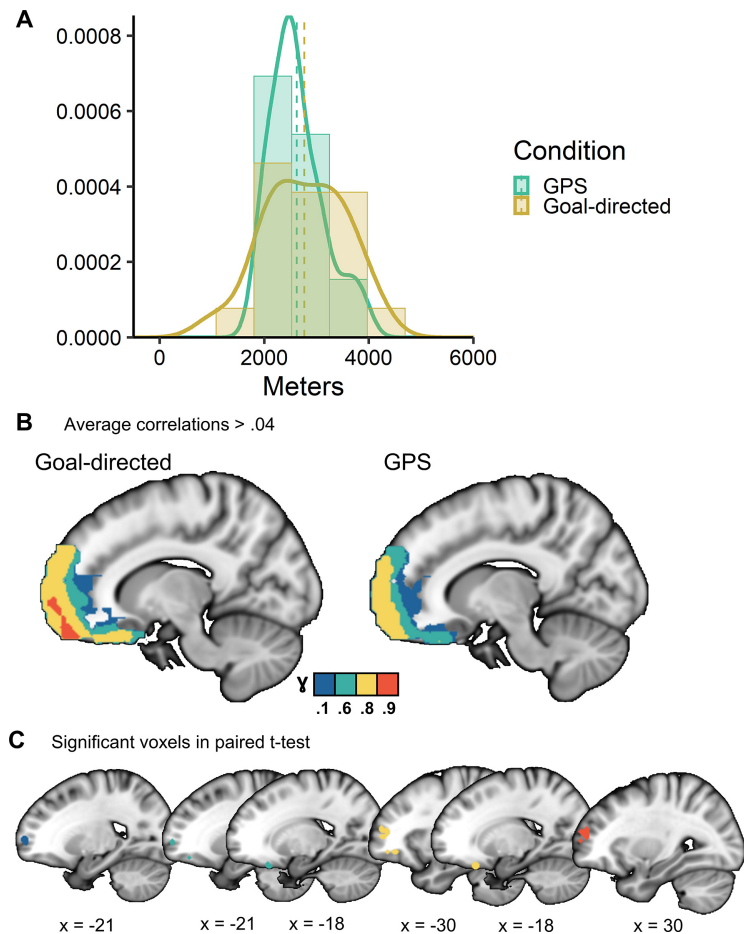

Figure 9. Predictive representations for goal-directed and GPS routes with matched distances. **A**, Distribution of distance-matched routes included in this analysis. **B**, Voxels with average correlation values of > 0.04. **C**, Significant voxels in goal-directed > GPS paired t test, thresholded at t value corresponding to 5% FPR. Colors represent predictive horizons corresponding to different discount parameters (blue: ɣ = 0.1; green: ɣ = 0.6; yellow: ɣ = 0.8; red: ɣ = 0.9).

directed condition. Pending replication with more controlled designs, these control analyses suggest that our main findings are reliable (Table 1; Figs. 7 and 8).

The second caveat: the selection of routes did not include multiple past and future trajectories for each state, nor multiple past routes for each goal location. Such designs would enable testing the graph structure of relations, advancing previous work using routes with multiple paths (Balaguer et al., 2016; Chanales et al., 2017), and dissociating pattern similarity because of recent memory from pattern similarity because of predictive representations more directly.

Follow-up fMRI studies can also investigate compressed representations and abstraction by asking whether states that appear on many paths have a pronounced predictive representation (e.g., subgoals, states with special graph properties), and whether nearby locations are clustered as one state (or subgoal) by some brain regions. Further studies could compare the temporal hierarchy of large-scale predictive representations for higher-level plans (e.g., train from New York to Philadelphia) and smaller subgoal processing (e.g., walk to the train station). One way to test this is to orthogonally manipulate distance and the number and location of subgoals, such as turns. Such designs could also test the dynamics of goal and subgoal representation, complementing existing electrophysiology and neuroimaging work showing goal representation in MTL and PFC (Howard et al., 2014; Brown et al., 2016; Sarel et al., 2017; Tsitsiklis et al., 2019).

Other interpretations

Theoretically, we have proposed that a given state has higher representational similarity to its frequently visited successors along the planned path (Ezzyat and Davachi, 2014; Momennejad et al., 2017). This can also be discussed in terms of increased association, integration, abstraction, and clustering (Ritvo et al., 2019); the spread of activation across memory networks (Sievers and Momennejad, 2019); or the replay of previous trajectories or paths (Wu and Foster, 2014; Ambrose et al., 2016; Momennejad et al., 2018). While there are clever analytic designs to hint one way or another, a clear-cut dissociation of these hypotheses requires higher spatiotemporal resolutions, such as electrophysiology and MEG.

In conclusion, we present support for the hypothesis that multiscale predictive representations in hippocampal-prefrontal hierarchies underlie cognitive maps and hierarchical planning. While posterior hippocampal regions supported smallest predictive scales, anterior prefrontal regions supported the largest predictive horizons. Follow-up studies can be designed to further investigate planning, subgoal setting, and abstraction in spatial and nonspatial tasks.

References

Ambrose RE, Pfeiffer BE, Foster DJ (2016) Reverse replay of hippocampal place cells is uniquely modulated by changing reward. Neuron 91:1124–1136.

Badre D, D'Esposito M (2007) Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. J Cogn Neurosci 19:2082–2099.

Balaguer J, Spiers H, Hassabis D, Summerfield C (2016) Neural mechanisms of hierarchical planning in a virtual subway network. Neuron 90:893–903.

Bartoń K (2020) MuMIn: MultiModel Inference. R package version 1.43.17.

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 67:1–48.

Behrens TE, Muller TH, Whittington JC, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018) What is a cognitive map? Organizing knowledge for flexible behavior. Neuron 100:490–509.

Bellmund JL, Gärdenfors P, Moser EI, Doeller CF (2018) Navigating cognition: spatial codes for human thinking. Science 362:eaat6766.

Ben-Shachar M, Lüdecke D, Makowski D (2020) effectsize: estimation of effect size indices and standardized parameters. J Stat Softw 5:2815.

Brown TI, Carr VA, LaRocque KF, Favila SE, Gordon AM, Bowles B, Bailenson JN, Wagner AD (2016) Prospective representation of navigational goals in the human hippocampus. Science 352:1323–1326.

Brunec IK, Bellana B, Ozubko JD, Man V, Robin J, Liu ZX, Grady C, Rosenbaum RS, Winocur G, Barense MD, Moscovitch M (2018) Multiple scales of representation along the hippocampal anteroposterior axis in humans. Curr Biol 28:2129–2135.e6.

Bunge SA, Kahn I, Wallis JD, Miller EK, Wagner AD (2003) Neural circuits subserving the retrieval and maintenance of abstract rules. J Neurophysiol 90:3419–3428.

Burgess N, Maguire EA, O'Keefe J (2002) The human hippocampus and spatial and episodic memory. Neuron 35:625–641.

Burgess PW (2000) Strategy application disorder: the role of the frontal lobes in human multitasking. Psychol Res 63:279–288.

Campbell K, Grigg O, Saverino C, Churchill N, Grady C (2013) Age differences in the intrinsic functional connectivity of default network subsystems. Front Aging Neurosci 5:73.

Chanales AJ, Oza A, Favila SE, Kuhl BA (2017) Overlap among spatial memories triggers repulsion of hippocampal representations. Curr Biol 27:2307–2317.e5.

Chen J, Hasson U, Honey CJ (2015) Processing timescales as an organizing principle for primate cortex. Neuron 88:244–246.

Christoff K, Gabrieli JD (2000) The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. Psychobiology 28:168–186.

Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, Holyoak KJ, Gabrieli JD (2001) Rostrolateral prefrontal cortex involvement in relational integration during reasoning. Neuroimage 14:1136–1149.

Christoff K, Keramatian K, Gordon AM, Smith R, Mädler B (2009) Prefrontal organization of cognitive control according to levels of abstraction. Brain Res 1286:94–105.

Cole MW, Etzel JA, Zacks JM, Schneider W, Braver TS (2011) Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. Front Hum Neurosci 5:142.

Collin SH, Milivojevic B, Doeller CF (2015) Memory hierarchies map onto the hippocampal long axis in humans. Nat Neurosci 18:1562–1564.

Constantinescu AO, O'Reilly JX, Behrens TE (2016) Organizing conceptual knowledge in humans with a gridlike code. Science 352:1464–1468.

Contreras M, Pelc T, Llofriu M, Weitzenfeld A, Fellous J (2018) The ventral hippocampus is involved in multi-goal obstacle-rich spatial navigation. Hippocampus 28:853–866.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Daw ND, Dayan P (2014) The algorithmic anatomy of model-based evaluation. Philos Trans R Soc Lond B Biol Sci 369:20130478.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215.

Dayan P (1993) Improving generalization for temporal difference learning: the successor representation. Neural Comp 5:613–624.

Duncan KD, Schlichting ML (2018) Hippocampal representations as a function of time, subregion, and brain state. Neurobiol Learn Mem 153:40–56.

Epstein RA, Patai EZ, Julian JB, Spiers HJ (2017) The cognitive map in humans: spatial navigation and beyond. Nat Neurosci 20:1504–1513.

Evensmoen HR, Lehn H, Xu J, Witter MP, Nadel L, Håberg AK (2013) The anterior hippocampus supports a coarse, global environmental representation and the posterior hippocampus supports fine-grained, local environmental representations. J Cogn Neurosci 25:1908–1925.

Ezzyat Y, Davachi L (2014) Similarity breeds proximity: pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. Neuron 81:1179–1189.

Fox J (2003) Effect displays in R for generalised linear models. J Stat Soft 8:1–27.

Fox J, Weisberg S (2019) An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage.

Garvert MM, Dolan RJ, Behrens TE (2017) A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. Elife 6:e17086.

Gilbert SJ (2011) Decoding the content of delayed intentions. J Neurosci 31:2888–2894.

Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523–534.

Hebscher M, Gilboa A (2016) A boost of confidence: the role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. Neuropsychologia 90:46–58.

Howard LR, Javadi AH, Yu Y, Mill RD, Morrison LC, Knight R, Loftus MM, Staskute L, Spiers HJ (2014) The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. Curr Biol 24:1331–1340.

Javadi AH, Emo B, Howard L, Zisch F, Yu Y, Knight R, Pinelo Silva J, Spiers H (2017) Hippocampal and prefrontal processing of network topology to simulate the future. Nat Commun 8:14652.

Jung M, Wiener S, McNaughton B (1994) Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. J Neurosci 14:7347–7356.

Keiflin R, Reese RM, Woods CA, Janak PH (2013) The orbitofrontal cortex as part of a hierarchical neural system mediating choice between two good options. J Neurosci 33:15989–15998.

Kjelstrup KB, Solstad T, Brun VH, Hafting T, Leutgeb S, Witter MP, Moser EI, Moser MB (2008) Finite scale of spatial representation in the hippocampus. Science 321:140–143.

Koechlin E (2011) Frontal pole function: what is specifically human? Trends Cogn Sci 15:241; author reply 243.

Koechlin E, Hyafil A (2007) Anterior prefrontal function and the limits of human decision-making. Science 318:594–598.

Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. Science 302:1181–1185.

Kuznetsova A, Brockhoff PB, Christensen RH (2017) lmerTest package: tests in linear mixed effects models. J Stat Softw 82:1–26.

Lee CS, Aly M, Baldassano C (2021) Anticipation of temporally structured events in the brain eLife 10:e64972.

Lohnas LJ, Duncan K, Doyle WK, Thesen T, Devinsky O, Davachi L (2018) Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus. Proc Natl Acad Sci USA 115:E7418–E7427.

McDannald MA, Takahashi YK, Lopatina N, Pietras BW, Jones JL, Schoenbaum G (2012) Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. Eur J Neurosci 35:991–996.

McDannald MA, Jones JL, Takahashi YK, Schoenbaum G (2014) Learning theory: a driving force in understanding orbitofrontal function. Neurobiol Learn Mem 108:22–27.

McKenzie S, Frank AJ, Kinsky NR, Porter B, Rivière PD, Eichenbaum H (2014) Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. Neuron 83:202–215.

Miller KJ, Botvinick MM, Brody CD (2017) Dorsal hippocampus contributes to model-based planning. Nat Neurosci 20:1269–1276.

Momennejad I (2020) Learning structures: predictive representations, replay, and generalization. Curr Opin Behav Sci 32:155–166.

Momennejad I, Haynes JD (2012) Human anterior prefrontal cortex encodes the 'what' and 'when' of future intentions. Neuroimage 61:139–148.

Momennejad I, Haynes JD (2013) Encoding of prospective tasks in the human prefrontal cortex under varying task loads. J Neurosci 33:17342–17349.

Momennejad I, Howard MW (2018) Predicting the future with multiscale successor representations. bioRxiv. doi: https://doi.org/10.1101/449470.

Momennejad I, Russek EM, Cheong JH, Botvinick MM, Daw ND, Gershman SJ (2017) The successor representation in human reinforcement learning. Nat Hum Behav 1:680–692.

Momennejad I, Otto AR, Daw ND, Norman KA (2018) Offline replay supports planning in human reinforcement learning. Elife 7:e32548.

Moscovitch M, Melo B (1997) Strategic retrieval and the frontal lobes: evidence from confabulation and amnesia. Neuropsychologia 35:1017–1034.

Nakagawa S, Johnson PC, Schielzeth H (2017) The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J R Soc Interface 14:20170213.

Nielson DM, Smith TA, Sreekumar V, Dennis S, Sederberg PB (2015) Human hippocampus represents space and time during retrieval of real-world memories. Proc Natl Acad Sci USA 112:11078–11083.

O'Keefe J, Nadel L (1978) The hippocampus as a cognitive map. Oxford: Clarendon.

Pauli WM, Gentile G, Collette S, Tyszka JM, O'Doherty JP (2019) Evidence for model-based encoding of Pavlovian contingencies in the human brain. Nat Commun 10:1099.

Peer M, Ron Y, Monsa R, Arzy S (2019) Processing of different spatial scales in the human brain. Elife 8:e47402.

Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE (2011) Statistical parametric mapping: the analysis of functional brain images. Elsevier: Amsterdam.

Poppenk J, Evensmoen HR, Moscovitch M, Nadel L (2013) Long-axis specialization of the human hippocampus. Trends Cogn Sci 17:230–240.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154.

Ramnani N, Owen AM (2004) Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat Rev Neurosci 5:184–194.

Ribas-Fernandes JJ, Shahnazian D, Holroyd CB, Botvinick MM (2019) Subgoal- and goal-related reward prediction errors in medial prefrontal cortex. J Cogn Neurosci 31:8–16.

Ritvo VJ, Turk-Browne NB, Norman KA (2019) Nonmonotonic plasticity: how memory retrieval drives learning. Trends Cogn Sci 23:726–742.

Ruediger S, Spirig D, Donato F, Caroni P (2012) Goal-oriented searching mediated by ventral hippocampus early in trial-and-error learning. Nat Neurosci 15:1563–1571.

Russek E, Momennejad I, Botvinick MM, Gershman SJ, Daw ND (2018) FMRI evidence for the successor representation in human value computation. Soc Neurosci Abstr.

Sarel A, Finkelstein A, Las L, Ulanovsky N (2017) Vectorial representation of spatial goals in the hippocampus of bats. Science 355:176–180.

Schapiro AC, Rogers TT, Cordova NI, Turk-Browne NB, Botvinick MM (2013) Neural representations of events arise from temporal community structure. Nat Neurosci 16:486–492.

Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM (2016) Statistical learning of temporal community structure in the hippocampus. Hippocampus 26:3–8.

Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA (2017) Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. Philos Trans R Soc Lond B Biol Sci 372:20160049.

Schlichting ML, Preston AR (2015) Memory integration: neural mechanisms and implications for behavior. Curr Opin Behav Sci 1:1–8.

Schlichting ML, Preston AR (2017) The Hippocampus and Memory Integration: Building Knowledge to Navigate Future Decisions. In: The Hippocampus from Cells to Systems (Hannula D, Duff M, eds). Springer, Cham.

Schlichting ML, Mumford JA, Preston AR (2015) Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat Commun 6:8151.

Schuck NW, Cai MB, Wilson RC, Niv Y (2016) Human orbitofrontal cortex represents a cognitive map of state space. Neuron 91:1402–1412.

Sievers B, Momennejad I (2019) SAMPL: the Spreading Activation and Memory PLasticity Model. bioRxiv. doi: https://doi.org/10.1101/778563.

Simon DA, Daw ND (2011) Neural correlates of forward planning in a spatial decision task in humans. J Neurosci 31:5526–5539.

Spalding KN, Schlichting ML, Zeithamova D, Preston AR, Tranel D, Duff MC, Warren DE (2018) Ventromedial prefrontal cortex is necessary for normal associative inference and memory integration. J Neurosci 38:3767–3775.

Spiers HJ, Gilbert SJ (2015) Solving the detour problem in navigation: a model of prefrontal and hippocampal interactions. Front Hum Neurosci 9:125.

Stachenfeld KL, Botvinick MM, Gershman SJ (2017) The hippocampus as a predictive map. Nat Neurosci 20:1643–1653.

Strange BA, Witter MP, Lein ES, Moser EI (2014) Functional organization of the hippocampal longitudinal axis. Nat Rev Neurosci 15:655–669.

Takahashi YK, Stalnaker TA, Roesch MR, Schoenbaum G (2017) Effects of inference on dopaminergic prediction errors depend on orbitofrontal processing. Behav Neurosci 131:127–134.

Tsitsiklis M, Miller J, Qasim SE, Inman CS, Gross RE, Willie JT, Smith EH, Sheth SA, Schevon CA, Sperling MR, Sharan A, Stein JM, Jacobs J (2019) Single-neuron representations of spatial memory targets in humans. bioRxiv. doi: https://doi.org/10.1101/523753.

van Kesteren MT, Beul SF, Takashima A, Henson RN, Ruiter DJ, Fernández G (2013) Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent. Neuropsychologia 51:2352–2359.

Vikbladh OM, Meager MR, King J, Blackmon K, Devinsky O, Shohamy D, Burgess N, Daw ND (2019) Hippocampal contributions to model-based planning and spatial memory. Neuron 102:683–693.e4.

Volle E, Gonen-Yaacovi G, de Lacy Costello A, Gilbert SJ, Burgess PW (2011) The role of rostral prefrontal cortex in prospective memory: a voxel-based lesion study. Neuropsychologia 49:2185–2198.

Whittington JC, Muller TH, Mark S, Chen G, Barry C, Burgess N, Behrens TE (2020) The Tolman-Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. Cell 183:1249–1263.

Wickham H (2016) ggplot2: elegant graphics for data analysis. New York: Springer.

Wikenheiser AM, Schoenbaum G (2016) Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. Nat Rev Neurosci 17:513–523.

Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. Neuron 81:267–279.

Wimmer GE, Büchel C (2019) Learning of distant state predictions by the orbitofrontal cortex in humans. Nat Commun 10:2554.

Wood JN, Grafman J (2003) Human prefrontal cortex: processing and representational perspectives. Nat Rev Neurosci 4:139–147.

Wu X, Foster DJ (2014) Hippocampal replay captures the unique topological structure of a novel environment. J Neurosci 34:6459–6469.

Yu LQ (2018) Neural mechanisms underlying schemas and inferences. J Neurosci 38:7930–7931.

Zeithamova D, Preston AR (2010) Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. J Neurosci 30:14676–14684.

Zeithamova D, Dominick AL, Preston AR (2012) Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. Neuron 75:168–179.