Predictive Representations in Hippocampal and Prefrontal Hierarchies

Iva K. Brunec[1], Ida Momennejad[2]
1 Department of Psychology, University of Toronto
2 Department of Biomedical Engineering, Columbia University

Abstract

As we navigate the world we learn about associations among events, extract relational structures, and store them in memory. This relational knowledge, in turn, enables generalization, inference, and hierarchical planning. Here we investigated relational knowledge during spatial navigation as multiscale predictive representations in the brain. We hypothesized that these representations are organized at multiple scales along posterior-anterior hierarchies in prefrontal and hippocampal regions. To test this, we conducted model-based representational similarity analyses of neuroimaging data measured during virtual reality navigation of familiar and unfamiliar paths with realistically long distances. We tested the pattern similarity of each point–along each navigational path–to a weighted sum of its successor points within different temporal horizons. Predictive similarity was significantly higher for familiar paths. Overall, anterior PFC showed predictive horizons at the largest scales (~875m) and posterior hippocampus at the lowest (~25m), while the anterior hippocampus (~175m), pre-polar PFC, and orbitofrontal regions (~350m) were in between. These findings support the idea that predictive representations are maintained at higher scales of abstraction in the anterior PFC, and unfolded at lower scales by pre-polar PFC and hippocampal regions. This representational hierarchy can support generalization, hierarchical planning, and subgoals at multiple scales.

Acknowledgements

The authors gratefully acknowledge Morris Moscovitch and Jason Ozubko for designing the original experiment, collecting and sharing the data, and sincerely thank Ken Norman, Buddhika Bellana, and Morgan Barense for helpful discussions. This work was supported by Canadian Institute of Health Research grant (#MOP49566 and #MOP125958) to Morris Moscovitch, a doctoral award from the Alzheimer Society of Canada (I.K.B.), grants from the James S. McDonnell Foundation and Natural Sciences and Engineering Research Council granted to Morgan Barense, NIMH Grant R01-MH104606 granted to Joshua Jacobs, and the John Templeton Foundation (I.M.), NIBIB R01EB022864.

Introduction

As we navigate the world, our brains construct and update representations in memory. During generalization and inference, relational structures are extracted from these representations. This relational knowledge is later retrieved to make decisions, plan, and guide behavior (Behrens et al., 2018). This idea has been captured by computational models that account for planning in human behavior (Momennejad et al., 2017), relational knowledge in human fMRI (Garvert, Dolan, & Behrens, 2017), and place and grid fields in rodent electrophysiology (Stachenfeld, Botvinick, & Gershman, 2017). This converging body of evidence suggests that the brain learns predictive maps of relational structures and uses them for fast and flexible planning. It has been suggested that these predictive representations are organized in a multi-scale fashion, each scale of representation corresponding to different gradients in the neural representational hierarchy, e.g., in the hippocampus (Momennejad & Howard, 2018; Stachenfeld et al., 2017) and the prefrontal cortex (Christoff & Gabrieli, 2000; Koechlin & Hyafil, 2007; Momennejad & Haynes, 2013). Here, we test the hypothesis that predictive representations are organized along prefrontal and hippocampal hierarchies at multiple scales.

Electrophysiological and neuroimaging evidence from rodents and humans indicate multi-scale representations of spatial locations in the hippocampus, or place fields, and entorhinal grid fields that tile the space with various levels of granularity (Brun et al., 2008; Kjelstrup et al., 2008; Poppenk, Evensmoen, Moscovitch, & Nadel, 2013; Strange, Witter, Lein, & Moser, 2014). Evidence from rodent electrophysiology suggests that the average place field size increases along the dorsoventral axis of the rodent hippocampus, with more ventral regions encoding space in a more overlapping manner and at a greater spatial scale (Contreras, Pelc, Llofriu, Weitzenfeld, & Fellous, 2018; Jung, Wiener, & McNaughton, 1994; Strange et al., 2014). Furthermore, human fMRI evidence suggests that the hippocampal posterior-anterior axis (homologous to the rodent dorsal-ventral axis) is also involved in representing memories from longer time-scales (Nielson, Smith, Sreekumar, Dennis, & Sederberg, 2015), finer-grained spatial representations (Evensmoen et al., 2013), and more abstract inference (Collin, Milivojevic, & Doeller, 2015).

The larger scale representations in the anterior hippocampus are proposed to support goal-directed search (Ruediger, Spirig, Donato, & Caroni, 2012), the integration of spatial and non-spatial states that are further apart (Collin et al., 2015), and longer time horizons (Brunec, Bellana et al., 2018; Nielson, Smith, Sreekumar, Dennis, & Sederberg, 2015); while those in the posterior hippocampus are more myopic and might support smaller predictive scales, such as smaller place fields (Strange et al., 2014) and pattern separation (Duncan & Schlichting, 2018; Leutgeb, Leutgeb, Moser, & Moser, 2007; Lohnas et al., 2018; Schlichting, Mumford, & Preston, 2015). Finally, recent computational models provide further support for the multi-scale property of predictive cognitive maps, explaining why place field representations are skewed toward goal

locations (Stachenfeld et al., 2017) and showing that multi-scale predictive representations can capture distance between states and reconstruct the sequence of experienced events (Momennejad & Howard, 2018). Taking these results into account, we hypothesized that the anterior hippocampus maintains predictive representations at a higher scale of abstraction compared to the posterior hippocampus.

Another candidate region for processing hierarchical representations during planning is the prefrontal cortex (Badre & D'Esposito, 2007). Broadly, it has been proposed that the prefrontal cortex (PFC) is involved in navigation when it is *active* and requires planning (Behrens et al., 2018; Epstein, Patai, Julian, & Spiers, 2017; Spiers & Gilbert, 2015), consideration of the number of paths to goal and alternative paths (Javadi et al., 2017), reversal and detours (Kringelbach & Rolls, 2004; Spiers & Gilbert, 2015), and retrospective revaluation mediated by offline replay (Momennejad, Otto, Daw, & Norman, 2018). Neuroimaging evidence suggests a prefrontal hierarchy in which more anterior PFC regions support relational reasoning (Christoff & Gabrieli, 2000; Christoff, Keramatian, Gordon, Smith, & Mädler, 2009), abstraction (Bunge, Kahn, Wallis, Miller, & Wagner, 2003; Christoff et al., 2001), and prospective memory (Gilbert, 2011; Haynes & Rees, 2006; Momennejad & Haynes, 2012, 2013). Bringing these findings together, we hypothesized that anterior prefrontal cortex encodes predictive maps with longer predictive horizons, i.e., information about states further away, while more posterior prefrontal regions maintain predictive maps with more myopic predictive horizons.

Importantly, we expected the scales of anterior prefrontal cortex to exceed the highest predictive horizons of the hippocampus (Figure 1A). This is because PFC cells broadly have longer delays enabling information to linger across longer time frames–as in working memory–and slower learning rates with the anterior PFC as the largest PFC region and a region in which we have the most difference with evolutionary ancestors (Ramnani & Owen, 2004). In contrast, the hippocampus has been suggested to be involved in rapid statistical learning at a faster rate (Schapiro, Turk-Browne, Botvinick, & Norman, 2017). It is important to note that the predictive horizon discussed here need not be merely temporal, or spatial, but can involve states that are far apart from one another in graph structures (Javadi et al., 2017) of relational knowledge acquired by statistical learning (Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013) and conceptual state space (Constantinescu, O'Reilly, & Behrens, 2016).

Here we tested the hypothesis that hippocampal-prefrontal hierarchies simultaneously maintain multiple predictive representations of the same underlying relational structures at different scales of abstraction. We used representational similarity analysis of an existing dataset with functional magnetic resonance imaging (fMRI) data. We reanalyzed fMRI data from a previously published study of realistic virtual navigation of familiar and novel routes (Brunec, Bellana et al., 2018). In this paradigm, participants underwent functional neuroimaging while they navigated familiar and unfamiliar routes in a virtualized version of the city they lived in (Toronto), built using images from Google Street View. This virtual navigation setup had important advantages for our purposes. First, it rendered the participants' experience as realistic as possible within the constraints of fMRI scanning. More importantly, it allowed us to compare pattern similarity for

long spatial horizons along distances used by humans in daily navigational planning (at the scale of kilometers). Finally, the experimental design benefited from participants' real world familiarity with certain paths, allowing us to compare the scales or predictive horizons of well-learnt long routes vs. novel routes. In order to navigate unfamiliar routes, participants used the same control buttons as the did along familiar routes, but they did not know the goal they were navigating towards and merely followed a dynamic arrow, akin to following a GPS (Figure 1B). These conditions will therefore be referred to as Familiar and GPS, respectively.

We used two main analyses to examine the structure of correlations between states along each trajectory (in Familiar and GPS conditions) to uncover multi-scale predictive representations along hippocampal and PFC hierarchies (Figure 1C). To maximally benefit from the level of resolution afforded by fMRI, paths were we split into steps such that every step corresponded to each time an entire brain volume was measured (repetition time, or TR). In the first analysis, we computed the correlation between every given step (TR) and the mean 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 the successor representation, we computed the correlation between every given time point and the weighted sum of future time points within a horizon. Each future TR was weighed exponentially using a discount parameter (i.e., gamma value, γ) between 0 and 1, and the value of the discount parameter corresponded to the scale of abstraction. Consistent with our prediction, we found that on Familiar, compared to GPS paths, the anterior hippocampus and anterior prefrontal regions maintained predictive maps with larger scales (i.e. relations to distal goals or successor states), while the posterior hippocampus cached predictive maps with smaller scales (i.e. relations to more proximal states).

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 due to 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). All participants had lived in Toronto for at least 2 years (M = 10.45, SE = 1.81). All participants were free of psychiatric and neurological conditions. All participants 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 upon completion of the study.

Experimental design and paradigm

We 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 v7.5.0.342. 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 4 conditions, and navigated 16 routes in total (4 in each condition, in a randomized order). The details of the experimental design have been reported in a previously published study (Brunec, Bellana, et al., 2018).

Data from two conditions of interest were analyzed in the present manuscript: Familiar and GPS/arrow-following routes. The routes were constructed prior to the date of scanning: participants built routes with researcher assistance, using a computer program which showed overhead maps of Toronto. Additionally, a set of 4 routes in areas of Toronto with which the participant reported no previous experience was created by the researcher to be used in the baseline (GPS) condition. In the scanner, participants were provided with Familiar route destinations and asked to navigate towards the goal along the most familiar/comfortable route. GPS trials involved no goal-directed navigation; instead, participants followed a dynamic arrow (Figure 1A). We only analyzed routes where participants successfully reached the goal ($M_{Familiar}$ = 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 visual input/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 echo-planar imaging (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 x 64 with a voxel size of 3.5mm x 3.5mm x 5.0mm. Images were first corrected for physiological motion using the Analysis of Functional NeuroImages (Cox, 1996).

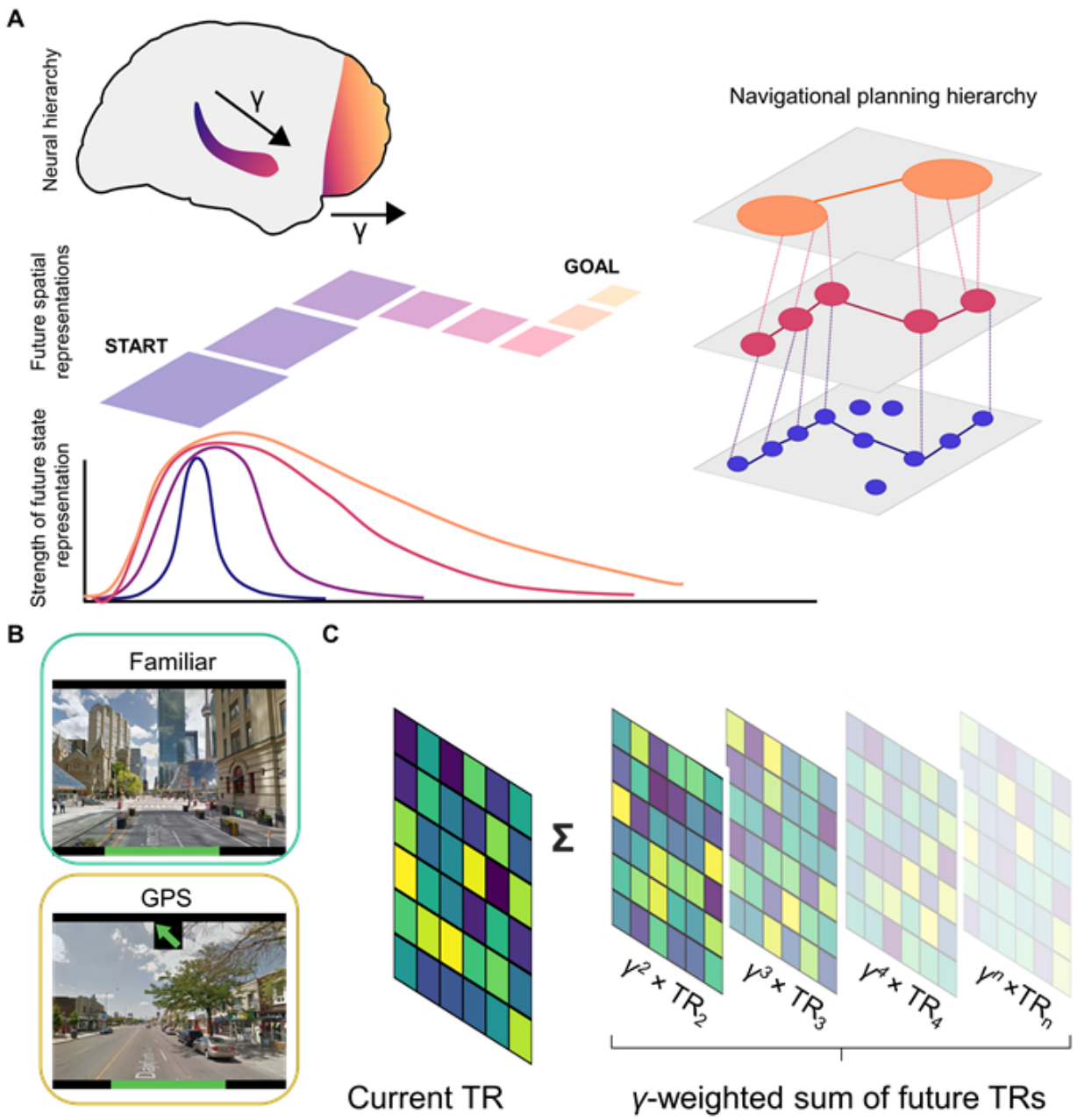


Figure 1. Schematic of the hypothesis, task conditions, and analytic methods. A) Predictive representations in the hippocampus and prefrontal cortex should proceed along a posterior-anterior gradient within the hippocampus and prefrontal cortex. Longer temporal horizons correspond to longer-range planning and greater scales of navigational representations. B) Participants used the same keys to navigate familiar routes and to follow the GPS dynamic arrow, but only familiar routes required goal-directed navigation. C) Analytic approach. The voxelwise pattern at each timepoint was correlated with the gamma-weighted sum of all future states (for gamma values of .1, .6, .8 and .9). With higher gamma values, the weighted future states remain above zero 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).

All subsequent preprocessing steps were conducted using the statistical parametric mapping software SPM12 (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011). Preprocessing involved slice timing correction, spatial realignment and co-registration with a resampled voxel size of 3 mm isotropic. No spatial smoothing was applied. The mean time-courses from participant-specific white matter and cerebrospinal fluid 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, Barnes, Snyder, Schlaggar, & Petersen, 2012), an additional motion scrubbing procedure was added to the end of our preprocessing pipeline (Campbell, Grigg, Saverino, Churchill, & Grady, 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 neighbouring data points. Motion scrubbing further minimizes any effects of motion-induced spikes on the BOLD signal, over and beyond standard motion regression, without leaving sharp discontinuities due to the removal of outlier volumes.

Analysis

Region of Interest analysis

We investigated the predictive similarity of each state to future representations in a set of regions of interest (ROIs). To do so, we first extracted voxelwise 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 timepoint (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 each timepoint (TR) within each navigated route with a weighted sum of future TRs. The future TRs were weighted by different constant values (γ), corresponding to different predictive spatial scales. The specified γ values were .1, .6, .8, and .9. With increasing γ values, timepoints further in the future remain weighted above zero.

As the average distance traversed within each TR was 25 meters, a γ value of .1 meant that only each subsequent step (1 TR away) was weighted above zero, and steps farther in the distance contributed little-to-no weight to the sum of future representations. For a γ value of .6, approximately 6-7 steps in the future were weighted above zero, corresponding to roughly 175 meters. For a value of .8, approximately 14 steps or 350 m were weighted above-zero, while this was the case for approximately 35 steps or 875 m for a γ value of .9.

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* ROI, including bilateral anterior and posterior hippocampi (aHPC, pHPC) and anterior and medial prefrontal cortical ROIs (antPFC, mPFC). As described in Brunec, Bellana et al. (2018), we divided the hippocampus into 6 anterior-posterior segments. We also examined the same measure in the dmPFC and antPFC. The anterior PFC and dorsomedial PFC 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 in (Russek, Momennejad, Botvinick, Gershman, & Daw, 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 z = -22, y = 56, z = -10). These analyses were performed for each of the ROIs, as well as a searchlight within the prefrontal cortex.

Prefrontal cortex searchlight analysis

In order to identify any gradients of predictive representation in the PFC, a custom searchlight analysis was performed within a prefrontal cortex mask (created in WFU PickAtlas). The analysis was restricted to grey matter voxels, and a spherical ROI with a 6mm radius was used to iteratively correlate each TR with the weighted sum of future states for voxels within each searchlight. The searchlight analysis was performed for four different values of ɣ: .1, .6., .8, and .9. The single-subject correlation maps were then compared against zero using the AFNI command 3dttest++. The output z-score maps were thresholded at values corresponding to 5% false positive rates.

Results

Participants navigated a set of distances they regularly traversed in everyday life ($M_{Familiar}$ = 4.5, M_{GPS} = 1.8 km). After completing each route, participants rated how familiar each route felt, and how difficult they found it to navigate on a scale from 1-9 (where 1 would correspond to most unfamiliar and most difficult, respectively). As expected, the average reported familiarity was higher in the Familiar condition (M = 7.0, SD = 1.79) than in the GPS condition (M = 3.0, SD = 1.06). The subjective difficulty was similar in the Familiar (M = 6.98, SD = 2.05) and GPS (M = 7.2, SD = 1.46) conditions, suggesting that all navigated routes were perceived to be similarly undemanding. The descriptive statistics for navigated distances, speed, and number of turns are displayed in Figure 2.

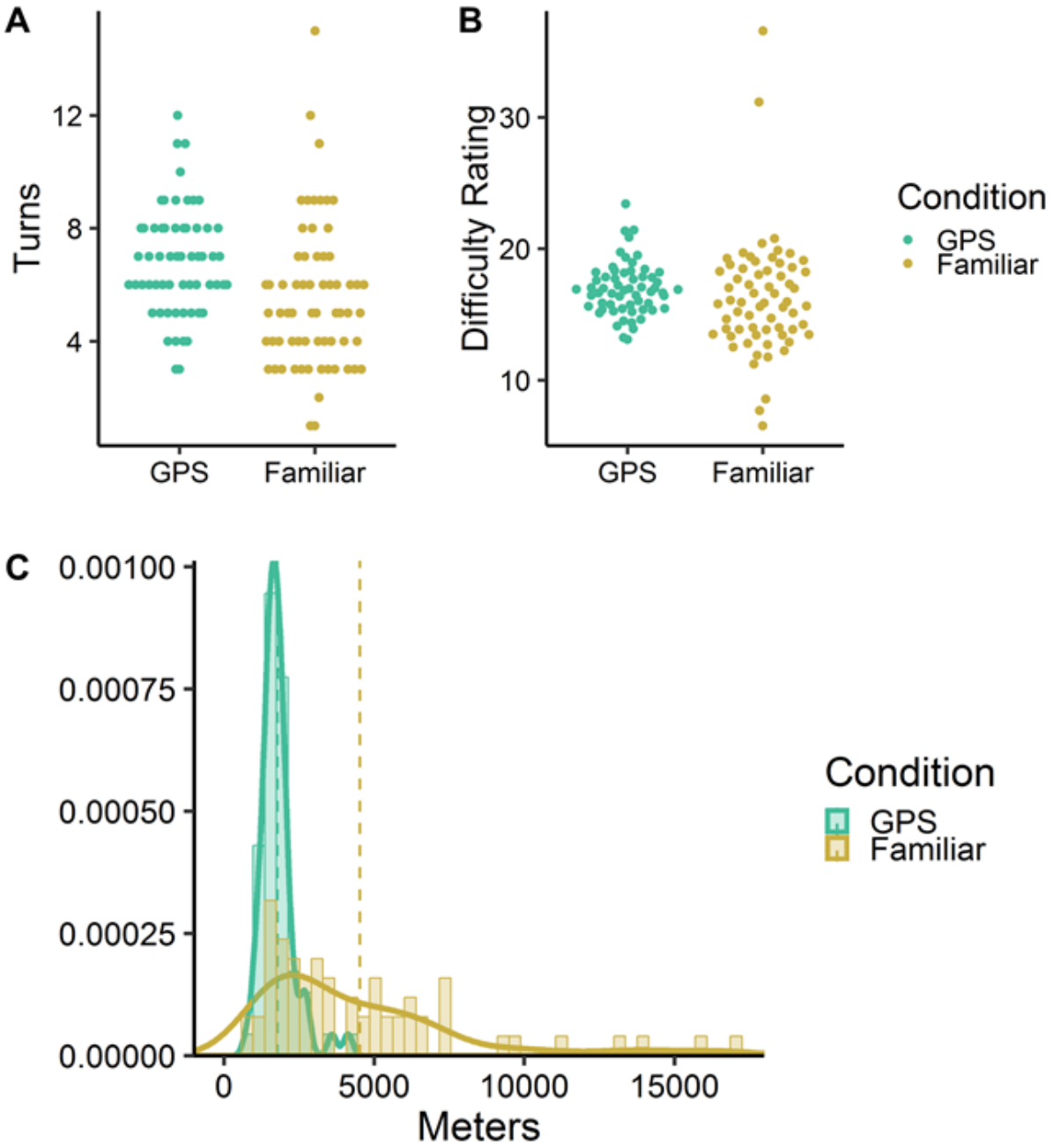


Figure 2. Descriptive statistics for navigated distances in Familiar and GPS conditions. A) The number of turns and B) the speed of travel was matched across the Familiar and GPS conditions. C) The distribution of travelled distances was highly different across the two conditions, with the average route being much shorter in the GPS, compared to the Familiar condition.

Hippocampal and prefrontal gradients of near-future predictive representations

To investigate predictive representations along the hippocampal longitudinal axis and within the PFC, we first performed an analysis investigating the similarity between each timepoint (TR) and the average of future 1, 2, 3, 4, 5, or 10 TRs. As described in Brunec, Bellana et al. (2018), we divided the hippocampus into 6 anterior-posterior segments. We also examined the same measure in the mPFC and antPFC.

We first ran linear mixed effects models on these similarity measures in bilateral hippocampi for each of the routes travelled within each condition, including the 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. 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, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 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. The similarity values for 10 TRs ahead were not entered in the present model due to the non-linear shift from 5 to 10 TR, but they are plotted in Figure 3.

We found a significant effect of axial segment ($F(5, 6796) = 45.38$, $p < .001$), driven by greater future representations in the anterior segments compared to posterior ones. There was also a main effect of condition ($F(1, 6796) = 1182.35$, $p < .001$), reflecting generally greater values in the Familiar (Figure 3A), compared to the GPS condition (Figure 3B), and a significant effect of the future horizon ($F(1, 6796) = 633.44$, $p < .001$), reflecting higher similarity values for states closer to the present. There was a main effect of hemisphere, reflecting higher values in the right compared to the left hemisphere ($F(1, 6796) = 6.97$, $p = .008$). There were significant interactions between axial segment and condition ($F(5, 6796) = 6.97$, $p < .001$), axial segment and future horizon ($F(20, 6796) = 2.13$, $p = .002$), and condition and future horizon ($F(4, 6796) = 13.16$, $p < .001$). The latter interaction is of particular interest as it suggests that the decline across different temporal horizons was greater in the GPS compared to the Familiar condition. There was no significant three-way interaction ($F < 1$).

We also ran the same models separately for the anterior and mPFC. In the antPFC, there was a significant main effect of condition ($F(1, 1222) = 363.76$, $p < .001$), as well as a main effect of future horizon ($F(4, 1222) = 48.36$, $p < .001$), but no condition by future horizon interaction ($F(4, 1222) = 1.18$, $p = .319$). In the mPFC, there was a significant effect of condition ($F(1, 1222) = 218.77$, $p < .001$) and a significant effect of future horizon ($F(4, 1222) = 114.82$, $p < .001$), but again no condition by future horizon interaction ($F(4, 1222) = 1.82$, $p = .122$).

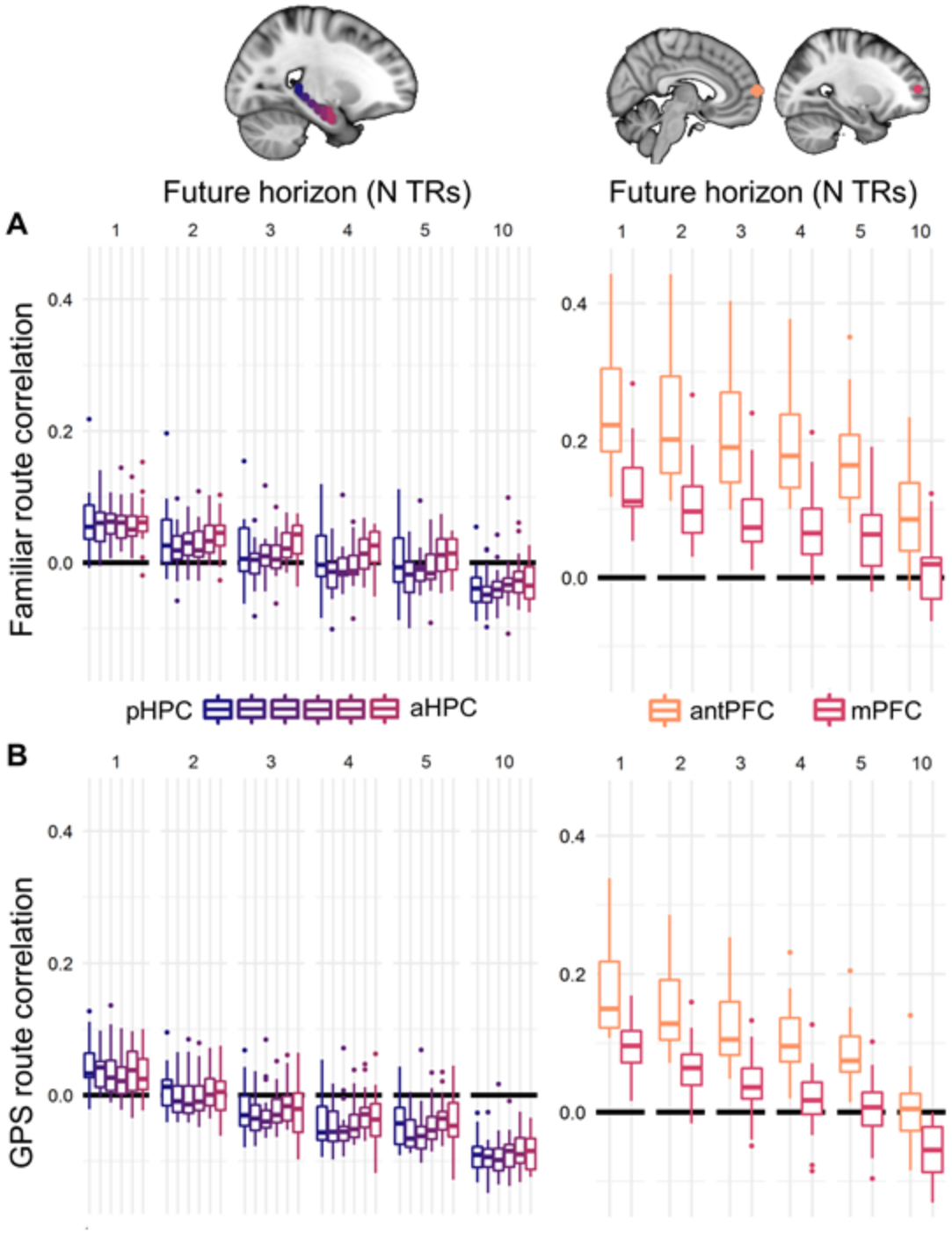


Figure 3. Similarity of each TR to mean of future TRs (equally weighted). Average correlation between each timepoint and the average of future 1/2/3/4/5 or 10 timepoints in A) the Familiar condition and B) the GPS condition. As the average distance traversed within each TR was 25 m (meters).

Comparing the representational similarity in the Familiar and GPS conditions against zero, we found that the anterior PFC displayed above-zero similarity for every predictive horizon including 10 steps ahead, in the Familiar condition (all p-values < .001), but only up to 5 steps in the GPS condition (all p-values for 1-5 steps < .001). In contrast, the medial PFC only displayed above-zero similarity up to 5 steps in the future on Familiar routes (p-values < .001) and three steps on GPS routes (p-values ≤ .002). The anterior-most hippocampal segment displayed above-zero similarity for up to 4 steps in the future (p-values ≤ .006) on Familiar routes and only one step on GPS routes (p < .001), while the posterior-most hippocampal segment displayed above-zero similarity for one step on Familiar routes (p < .001), and two steps on GPS routes (p-values ≤ .006).

Model-based (weighted sum) predictive representations in ROIs

To investigate the similarity between each time point and ɣ-weighted representations of future states, we again 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 gamma and condition as fixed effects and participant as a random effect. For the hippocampus, the reported statistics and plotted values apply to the right hippocampus, but there was no significant difference between the left and right hippocampi (all ps > .34).

The first mixed effects model included all ROIs to determine whether the average correlation values differed across regions with different hypothesized future timescales. There was a significant main effect of ɣ, suggesting that all regions showed a decrease in correlation with increasing values of ɣ ($F(1, 1456) = 565.22$, $p < .001$), as well as a significant main effect of condition, suggesting that correlations were generally greater in the Familiar (Figure 4A), compared to the GPS condition ($F(1, 1457) = 31.99$, $p < .001$; Figure 4B). There was a main effect of ROI ($F(3, 1456) = 151.77$, $p < .001$), confirming the prediction of strongest future representations in the antPFC, followed by mPFC, aHPC, and pHPC. There was also a significant interaction between ɣ and condition ($F(1, 1456) = 13.97$, $p < .001$), and a non-significant, trending interaction between condition and ROI ($F(3, 1456) = 2.57$, $p = .053$).

Follow-up mixed effects models were run for values *within* each ROI and tested against a Bonferroni-adjusted value of α = .0125 (as 4 ROIs were investigated). In the antPFC, there was a significant main effect of ɣ, with significantly higher correlations for lower values of gamma ($F(1, 349) = 86.78$, $p < .001$). There was also a significant effect of condition, with significant higher correlations in the Familiar than the GPS condition ($F(1, 350) = 16.37$, $p < .001$). There was no significant ɣ x condition interaction ($F(1, 349) = 1.52$, $p = .218$).

In mPFC, there was again a significant main effect of ɣ ($F(1, 352) = 182.29$), as well as a main effect of condition ($F(1, 353) = 5.83$, $p = .016$) in the same direction as the antPFC. There was also a significant ɣ x condition interaction ($F(1, 352) = 6.23$, $p = .011$). This interaction reflected

a steeper slope across the different values of γ in the GPS condition (slope = -.18), compared to the values in the Familiar condition (slope = -.12), although γ was significant in both conditions (both p-values < .001).

In the aHPC, there was a significant main effect of γ ($F(1, 350) = 267.48$, $p < .001$), a main effect of condition ($F(1, 350) = 9.48$, $p = .002$), as well as a γ x condition interaction ($F(1, 350) = 8.94$, $p = .003$). As in the mPFC, this interaction reflected a steeper slope across γ values in the GPS condition (-.16) than in the Familiar condition (-.12).
In the pHPC, there was a significant main effect of γ ($F(1, 351) = 398.97$, $p < .001$), a main effect of condition ($F(1, 351) = 6.01$, $p = .015$), and a significant γ x condition interaction ($F(1, 351) = 6.95$, $p = .009$), again reflecting a steeper slope in the GPS condition (-.17), compared to the Familiar condition (-.13).

Because the conditions differed in the range of distances participants traversed, we also ran the same set of analyses with distance as a covariate. These are reported in Supplemental Materials, but the significance of the main effects and interactions reported above remained largely unchanged, with the exception of the main effects of condition in the mPFC and pHPC. This suggests that representations of future states in the antPFC and aHPC are driven by goal-directed navigation, and not the objective distance traversed.

To test for evidence of predictive representations, we tested these values against zero, with an adjusted value of $\alpha = .002$ (24 comparisons in total). At $\gamma = .1$, the correlations in all ROIs were significantly above zero in both conditions. At $\gamma = .6$, all correlations in the Familiar condition were significantly above zero, with the exception of pHPC. In the GPS condition, however, correlations in neither the aHPC nor the pHPC were significantly above zero. At $\gamma = .8$, values in both antPFC and mPFC remained significantly above zero in the Familiar condition, but only antPFC remained above zero in the GPS condition. For this value of γ, the values in aHPC and pHPC were not significantly above zero in either condition, and were actually significantly below zero 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.

Controlling for Distance in ROI Analysis

To control for different distances travelled by the participants in the Familiar vs. GPS conditions, we ran the same set of analyses as reported above, but added the distance covered on each route as a covariate to our mixed models. The distances were mean-centered within each participant.

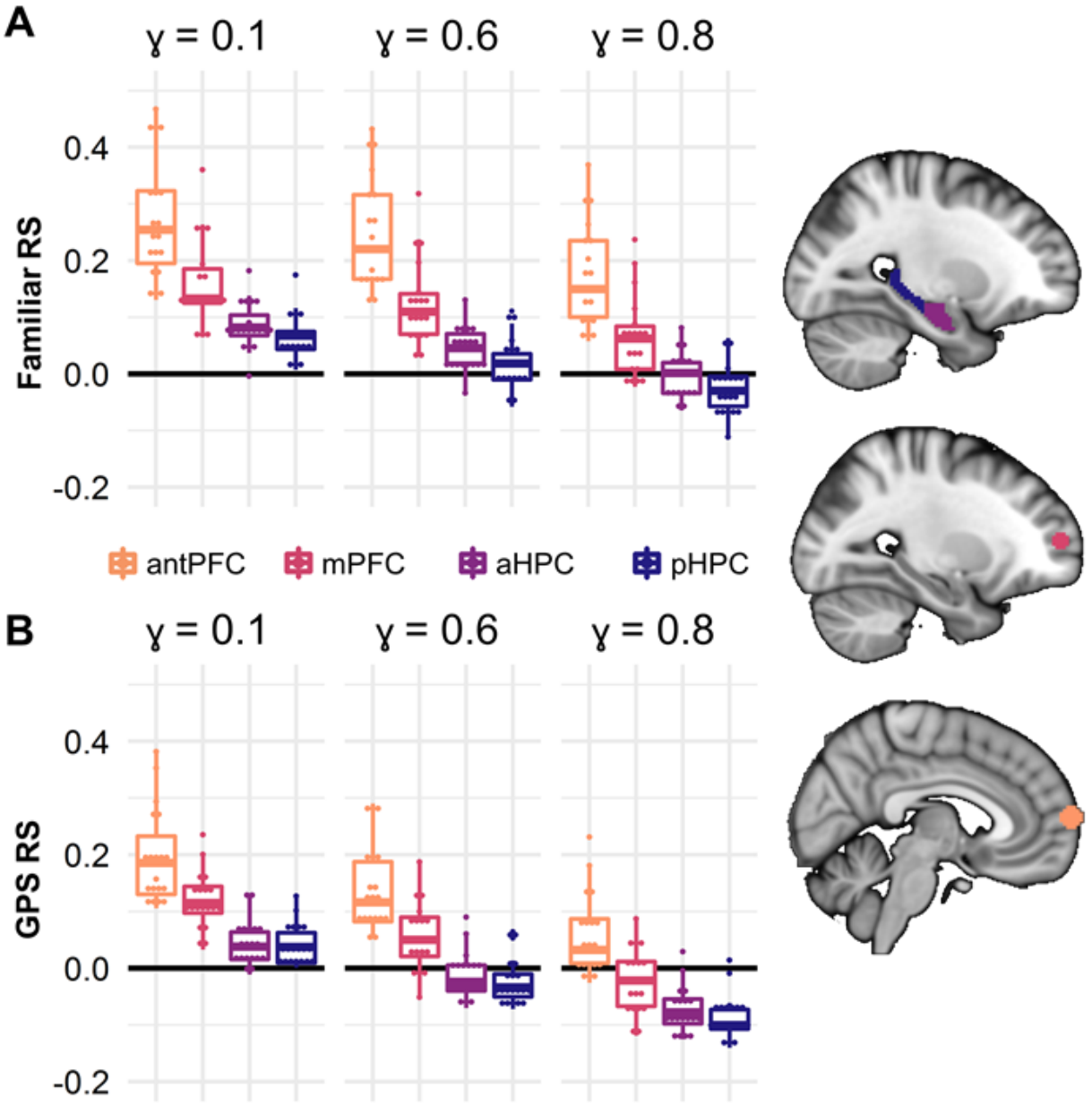


Figure 4. Predictive similarity across the scales. Correlations between current timepoints and the ɣ-weighted sum of future states for different values of gamma, in the four specified ROIs in the A) Familiar and B) GPS conditions. ɣ=.1 only included 1-step (1 TR) away, ɣ=.6 reached approximately 6-7 steps in the future, corresponding to roughly 175 m, ɣ=.8, approximately 14 steps or 350 m ahead.

In the overall model, there was a significant main effect of distance, with greater distances corresponding to generally higher correlation values ($F(1, 1456) = 16.78$, $p < .001$). However, the addition of this covariate did not change the significance of the main effects of ɣ, condition, ROI, nor the ɣ x condition interaction (all p-values < .001).

The same pattern emerged in the follow-up models run within each of the four ROIs. In the antPFC, there was again a significant main effect of distance ($F(1, 349) = 6.02$, $p = .015$), but this did not change the significance of the main effects of γ or condition (respective p-values < .001 and .002). In the mPFC, the effect of distance on correlations was non-significant, but trending ($F(1, 352) = 3.27$, $p = .071$), but the main effect of γ remained significant ($p < .001$), as did the γ x condition interaction ($p = .014$). However, the main effect of condition was no longer significant after accounting for distance within this ROI ($p = .076$).

In the aHPC, distance was again a significant covariate ($F(1, 350) = 5.62$, $p = .018$), but this did not change the significance of γ ($p < .001$), condition ($p = .027$), or the γ x condition interaction ($p = .003$). In the pHPC, distance was also a significant covariate ($F(1, 350) = 9.47$, $p = .002$), and while the significant main effect of γ ($p < .001$) and the γ x condition interaction ($p = .009$) remained significant, the effect of condition did not ($p = .145$).

Together, these analyses suggest that while distance travelled accounted for a significant amount of variance in the correlation data, it largely did not confound the effects of the factors of γ and condition. It is worth noting that a significant effect of condition persisted aHPC and antPFC, but not the pHPC and mPFC after adding distance travelled to the model, suggesting that correlations with future states in the antPFC and aHPC are driven by goal-directed navigation, and not the objective distance traversed.

Predictive representations in prefrontal searchlights

Prefrontal cortex has a much larger volume than the hippocampus. In order to identify hierarchies of predictive representations comparable to hippocampal ROIs, we ran a searchlight analysis and computed similarity for voxels within every spherical searchlight (of 6mm radius). The searchlight analysis was performed for four values of γ (.1, .6, .8., .9) within each of the conditions. The thresholded z-score maps for different values of γ are displayed as overlays in Figure 5A, along with the average correlation maps within each condition (thresholded at .06; Figure 5B).

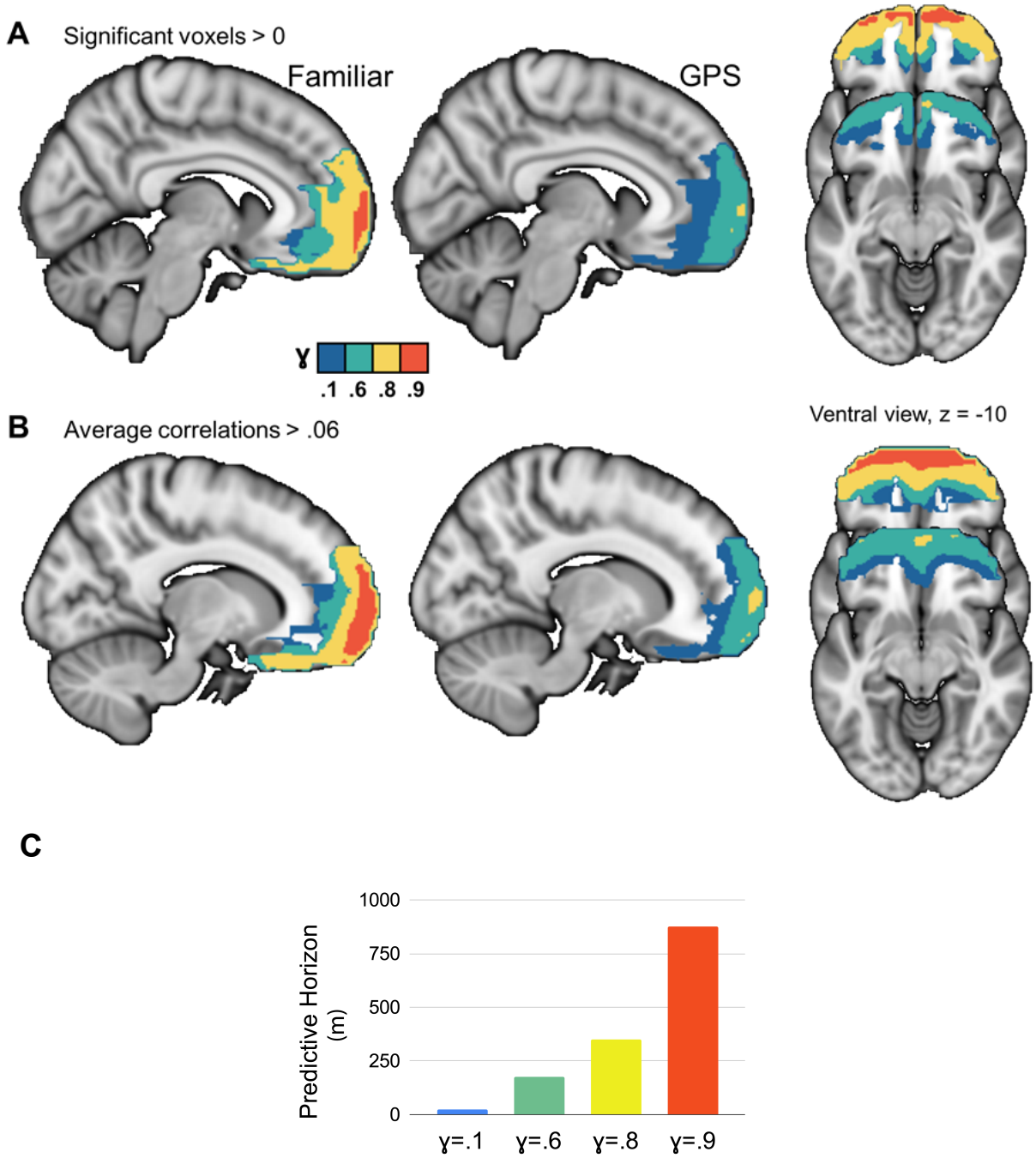


Figure 5. One-sample T-tests for Familiar and GPS condition. A) Voxels with significant representations of future states in the Familiar and GPS conditions using a one-sample t-test against zero. B) Voxels with representational similarity (correlation) values above .06 for each value of ɣ. *See Supplementary Figure 2 for the same results controlled for distance by adding it as a covariate within each condition*. C) Distances corresponding to the predictive horizons corresponding to each discount parameter. ɣ=.1 only included 1-step (1 TR) away, ɣ=.6 reached approximately 6-7 steps in the future, corresponding to roughly 175 m, ɣ=.8, approximately 14

steps or 350 m, ɣ=.9 reached approximately 35 steps or 875 m ahead. (Blue: ɣ=.1, Green: ɣ=.6, Yellow: ɣ=.8, Red: ɣ=.9)

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 6. These plots reveal a gradation of future state representations extending from posterior- to anterior-most slices of the PFC. This trend was reliable in both the Familiar and GPS conditions, but the representational similarity values were consistently greater in the Familiar condition.

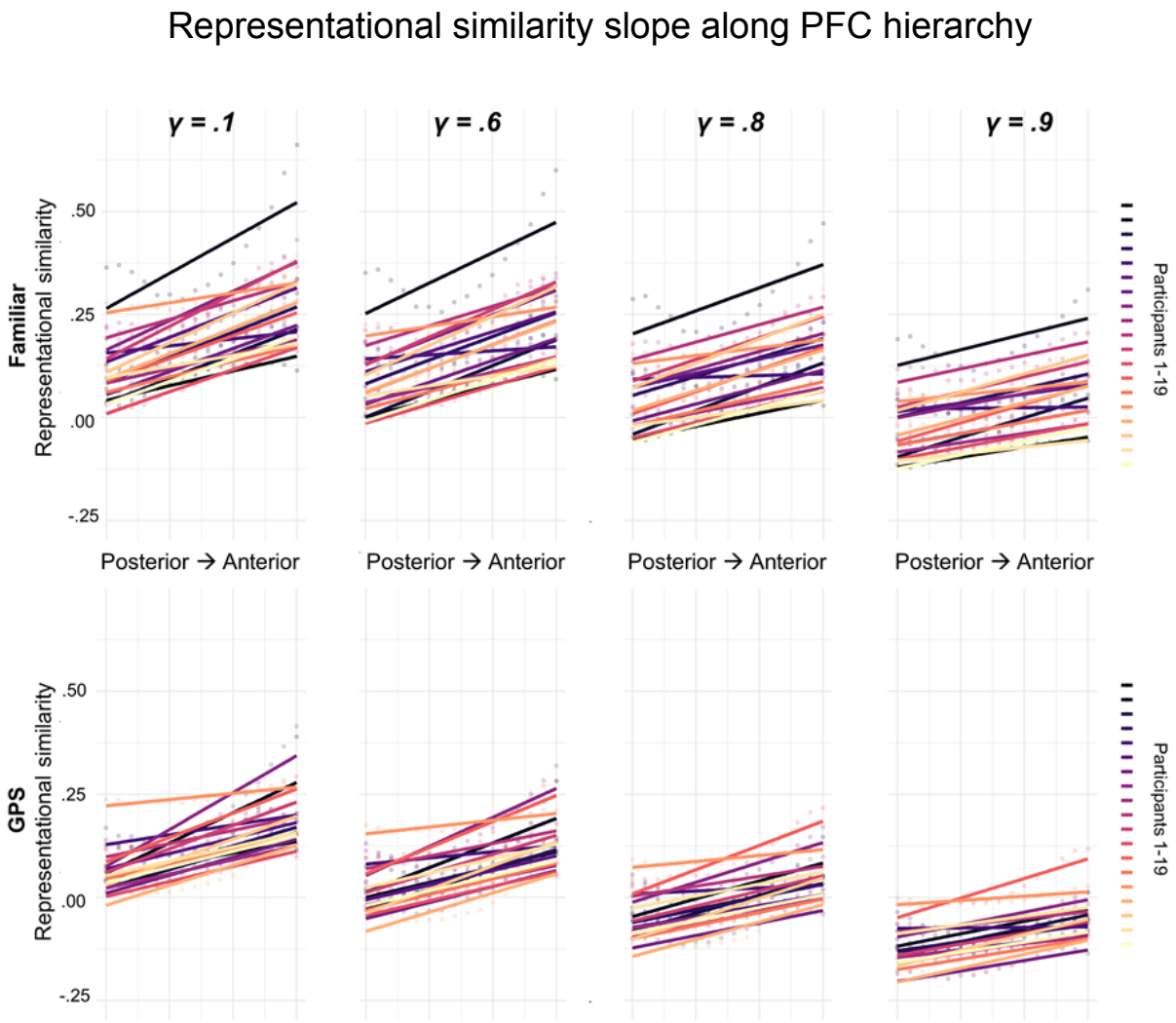


Figure 6. Increasing predictive similarity along posterior-to-anterior PFC. In order to indicate which PFC regions displayed higher predictive similarity we computed the slope of correlations for posterior-to-anterior PFC slices for Familiar and GPS conditions. We computed these slopes for four values of ɣ, corresponding to gradients of low to high scales. Each line corresponds to predictive similarity results from one of 19 participants.

To account for the proportion of different histologically-defined brain regions covered by each significant cluster, we calculated the % of overlap between each prefrontal Brodmann Area (BA) region and the significant voxels for each value of ɣ in each of the conditions. These percentages are reported in Table 1 and Figure 7. These percentages represent the proportion of each BA region covered by the significant thresholded clusters. We found the largest overlap between voxels in the anterior PFC (BA 10) and significant voxels in the searchlight analysis with various ɣ values. Following anterior and polar PFC was BA 11, corresponding to the orbitofrontal cortex, and then BA 25 and 32, corresponding to subgenual area or cingulate cortex and anterior cingulate cortex respectively. These regions were followed by smaller overlap in area 47, corresponding to the orbital part of the inferior frontal gyrus, areas 46and 9 corresponding to the dorsolateral PFC, and no overlap in area 45 corresponding to the inferior frontal gyrus.

	Familiar				**GPS**			
	ɣ = .1	**ɣ = .6**	**ɣ = .8**	**ɣ = .9**	**ɣ = .1**	**ɣ = .6**	**ɣ = .8**	**ɣ = .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%

Table 1. Proportion of each prefrontal Brodmann Area accounted for by the significant prefrontal voxels. Results were driven from the one-sample T-test results displayed in Figure 5.a. (not matched for distance). Proportions are displayed for each value of ɣ within each condition.

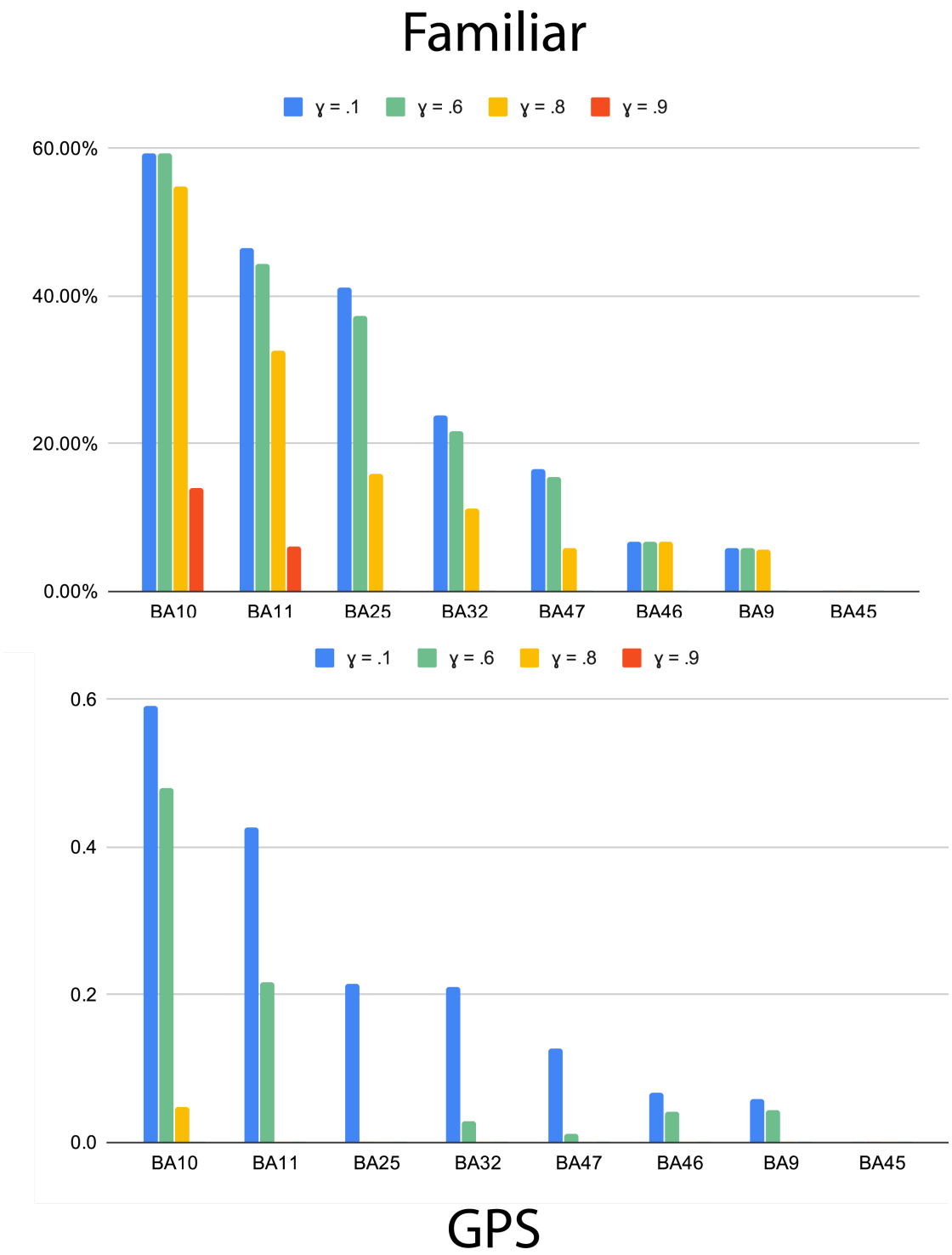


Figure 7. Prefrontal cortex hierarchy in the Familiar and GPS conditions. Proportion of prefrontal Brodmann Areas accounted for by the significant PFC voxels in searchlight analysis are shown. Results were driven from the one-sample T-test results displayed in Figure 5.a. (not matched for distance). Colorbars reflect different discount values (γ) corresponding to different predictive horizons within each condition. (Blue: γ=.1, Green: γ=.6, Yellow: γ=.8, Red: γ=.9).

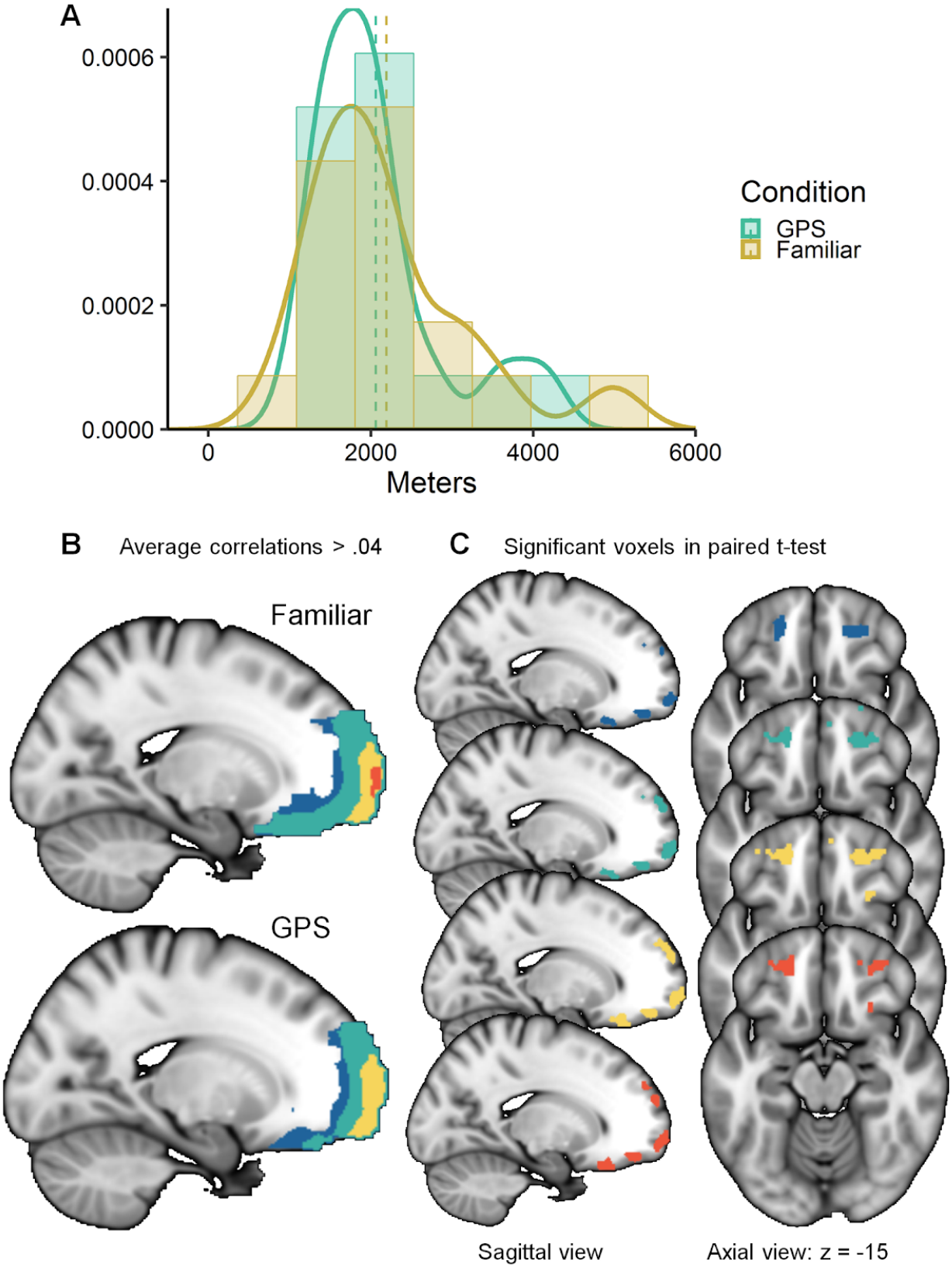

A
0.0006
0.0004
0.0002
0.0000
0
2000
4000
6000
Meters
Condition
GPS
Familiar
B
Average correlations > .04
Familiar
GPS
C
Significant voxels in paired t-test
Sagittal view
Axial view: z = -15

Figure 8. Predictive representations for familiar and GPS routes with matched distances. A) Distribution of distance-matched routes included in this analysis. B) Voxels with average correlation values of > .04. C) Significant voxels in Familiar > GPS paired t-test, Bonferroni corrected. Colors reflect predictive horizons corresponding to different discount parameters (Blue: ɣ=.1, Green: ɣ=.6, Yellow: ɣ=.8, Red: ɣ=.9).

Controlling for Distance: Matched Distance Analysis

As discussed in the ROI sections, the distances were not matched between the two conditions (Figure 2C). 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 a kilometer. We were unable to include 3 of the participants in this analysis as the distances in their Familiar and GPS routes were too different (with distance differences of 1.5, 2.7, and 3.7km). For the remaining 16 participants, there was no significant difference between the selected GPS and Familiar routes ($p = .180$). We ran a paired samples t-test comparing their prefrontal correlation maps for the two selected routes. The brain maps of the average correlation values thresholded at .06 are presented in Fig. 8A and the results of the Bonferroni-corrected t-test in Fig. 8B.

We compared matched-distance searchlight results in the Familiar and GPS conditions. In this comparison, relatively few clusters significantly differed between the Familiar and GPS conditions. However, the comparison at each level of ɣ suggests that there is a reliable 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 familiar and GPS conditions. It is worth noting, however, to ensure matched distances between the Familiar 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, likely resulting in greater noise and lower statistical power.

Discussion

We investigated the hypothesis that relational knowledge–about navigational paths–is organized as multi-scale predictive representations in hippocampal and prefrontal hierarchies. In a virtual navigation task where participants navigated familiar and unfamiliar paths with realistically long distances, we analyzed the representational similarity of each location on the path with locations along a prospective horizon. Motivated by previous work on multi-scale predictive representations (Momennejad & Howard, 2018), we computed pattern similarity to discounted predictive horizons into the future (25-875 meters)–i.e., sum of TRs weighed by a discount

parameter into the future. These analyses revealed four main findings. First, predictive similarity were of longer horizons in familiar paths compared to non-familiar paths. Second, similarity in anterior hippocampus and frontopolar cortex were significantly higher in the Familiar condition and for longer horizons. Third, predictive representations were organized along a posterior-anterior hierarchy of predictive horizons (25-175m) in the hippocampus with larger scales in gradually more anterior hippocampal regions. Fourth, similarity to future horizons was organized along a rostro-caudal prefrontal hierarchy along the orbitofrontal cortex, medial and lateral PFC, and anterior PFC with larger scales horizons (25-875m) in gradually more polar regions (Figure 5). Overall, anterior PFC showed predictive similarity at the largest scales and posterior hippocampus the lowest, while the anterior hippocampus, pre-polar PFC, and orbitofrontal regions were in between.

These results support the hypothesis that prefrontal-hippocampal interactions organize relational knowledge–in this case for spatial navigation–at different scales of generalization and abstraction (Behrens et al., 2018; Momennejad & Howard, 2018). In the case of spatial navigation, this hierarchical representation enables hierarchical planning and subgoal computation using graphs of the environment (Figure 1) abstracted at different scales (Ribas-Fernandes, Shahnazian, Holroyd, & Botvinick, 2018). Our proposal is that planning at larger scales may be enabled by larger and more abstract scales of representations in anterior PFC (Figure 1, large scale graph). This higher level plan may be translated into more precise policies using representations in pre-polar PFC and anterior hippocampal regions (Figure 1, mid-scale graph), and finer scale trajectories are translated by hippocampal gradients down to the smallest predictive horizons of place fields (Figure 1, small scale graph). This proposal is also supported by previous findings.

Consistent with our proposal, recent work on cognitive maps in rodents, monkeys, and humans indicate PFC's involvement in *active* navigation and planning (Epstein et al., 2017), while earlier work on finer scale spatial representation had primarily focused on the hippocampus. The hippocampus is thought to support cognitive maps of space (O'Keefe & Nadel, 1978; Burgess, Maguire, & O'Keefe, 2002) as well as nonspatial relational structures (Bellmund, Gärdenfors, Moser, & Doeller, 2018; Garvert et al., 2017). Recent computational perspectives suggest that the hippocampus serves rapid statistical learning (Schapiro et al., 2017; Schapiro, Turk-Browne, Norman, & Botvinick, 2016) to form and update a predictive map of the state space at multiple scales (Momennejad & Howard, 2018; Stachenfeld et al., 2017). As such, the hippocampus serves as a predictive map that organizes relational knowledge of spatial and non-spatial states (Garvert et al., 2017; McKenzie et al., 2014; Schuck, Cai, Wilson, & Niv, 2016). Notably, the long axis of the hippocampus is shown to support gradually larger spatiotemporal scales (Brunec et al., 2018; Nielson et al., 2015; Poppenk et al., 2013; Strange et al., 2014). Place fields have been especially well-studied in the dorsal-ventral rodent hippocampal axis (Strange et al., 2014), while a few fMRI studies have focused on the role of posterior-anterior axis in spatio-temporal scales (Nielson et al., 2015), and inference on mnemonic relations (Collin et al., 2015; Schlichting & Preston, 2015).

An important aspect of our findings is that we observed predictive similarity with gradually higher scales along the posterior-to-anterior gradient of the PFC hierarchy (Figures 5, 6, 7, 8, Table 1). To compare predictive similarity in this gradient, we computed the slope of correlations for (weighted sum of TRs within) each predictive horizon across posterior to anterior PFC slices. We found an overall effect of condition, where predictive similarity was generally higher in the Familiar vs. the GPS condition especially for higher horizons, as well as a prefrontal gradient effect: more anterior PFC regions showed higher predictive similarity (correlation) values in general. Furthermore, we measured the proportional overlap between significant voxels in the searchlight analysis and voxels in different histologically-defined PFC regions. To do so, we calculated the % of overlap between each prefrontal Brodmann Area (BA) region and the significant voxels for each value of γ in the familiar and GPS conditions. These percentages, reported in Table 1 and Figure 7, clearly indicate the largest overlaps in the anterior PFC (BA 10), orbitofrontal cortex (BA 11), and granular and anterior cingulate cortex (BA 25 and 32). These findings are consistent with the direction of the slope of predictive similarity in Figure 6. Here we contextualize the PFC findings reported here in light of previous studies of cognitive maps that have focused on the role of the prefrontal cortex.

When we compared the proportion of voxels within PFC subregions contributing to predictive similarity, we found that the polar or anterior PFC–largely corresponding to Brodmann area 10 (BA 10)– had the highest proportion of voxel overlap, especially for larger predictive scales (Table 1). BA 10 is the largest cytoarchitectonic region of the human PFC, it has the largest volumetric and proportional difference between humans and other great apes, it is highly interconnected within the PFC, and its cells display longer decay times (Ramnani & Owen, 2004). Thus, the properties of BA 10 suggest a structurally well-connected region to support higher levels of abstraction, e.g., predictive representations with larger scales of integration, which can be thought of in terms of clustering of relational graphs with a higher radius. For temporal relations, this graph clustering or integration radius can be thought of in terms of longer decays or longer sustained memory leading to binding over longer time-scales. For spatial relations, this radius can be thought of in terms of associating locations that are further apart. For relational structures, this radius can be thought of in terms of an increase in similarity among a cluster of associations within a given degree of separation.

A related–and crucially non-spatial–body of evidence from the study of goal-directed behavior and prospective memory indicates a functional role for anterior or rostral prefrontal cortex (also known as the frontopolar cortex) in the encoding and retrieval of prospective task sets and targets (Gilbert, 2011; Haynes & Rees, 2006; Momennejad & Haynes, 2012, 2013). This frontopolar evidence fits well with the proposal that the PFC is organized in a rostrocaudal hierarchy (Badre & D'Esposito, 2007; Koechlin, 2011; Koechlin & Hyafil, 2007; Koechlin, Ody, & Kouneiher, 2003), with more anterior or rostral regions corresponding to higher levels of abstraction and relational (Bunge et al., 2003; Kalina Christoff et al., 2009; Ida Momennejad & Haynes, 2013).

Studies and models of the orbitofrontal cortex (OFC) have indicated the OFC as the brain's cognitive map of task-related state spaces that serves prediction, decision-making, and planning (Schuck et al., 2016; Wilson, Takahashi, Schoenbaum, & Niv, 2014). However, some studies suggest that the involvement of the OFC is more tied with the anticipation of reward (Kahnt, Heinzle, Park, & Haynes, 2010), reversal learning and reappraisal due to prediction errors (Boorman, Behrens, Woolrich, & Rushworth, 2009), and prediction of states-value associations (Wimmer & Büchel, 2019). One interpretation of these findings has been that since ventral PFC regions including the OFC are well-connected to the subcortical value system, such as the ventral striatum, they are functionally well suited to process state space representations in connection to value; while dorsal PFC is better connected to the dorsal striatum and motor cortical regions and is therefore better suited to manage action policies.

Furthermore, OFC as well as anterior PFC have been suggested to support model-based reinforcement learning (Daw, Gershman, Seymour, Dayan, & Dolan, 2011), where an animal unfolds a learned state-action-state associative model during planning and decision making. This finding that has been replicated across different experiments (Daw & Dayan, 2014; McDannald, Jones, Takahashi, & Schoenbaum, 2014; McDannald et al., 2012; Pauli, Gentile, Collette, Tyszka, & O'Doherty, 2019), and supports the idea that the OFC maintains state-state relational maps that support iterative value computation in planning and decision-making (Daw, Niv, & Dayan, 2005; Keiflin, Reese, Woods, & Janak, 2013; Simon & Daw, 2011). Notably, recent work on the neural substrates of model-based behavior indicate a role for the hippocampus in model-based decision-making as well (Miller, Botvinick, & Brody, 2017; Vikbladh et al., 2019). Consistently in our findings, predictive representations in the anterior hippocampus were the most similar to OFC representations, however, more anterior OFC regions yielded predictive similarity within horizons of higher predictive scales than the hippocampus (Figures 4, 8). These findings expand previous perspectives on OFC-hippocampal interactions in cognitive map like representations (Keiflin et al., 2013; Schuck et al., 2016; Wikenheiser & Schoenbaum, 2016; Wood & Grafman, 2003).

While the present analyses were focused on spatial navigation, predictive representations are generalizable to non-spatial domains such as relational knowledge and category generalization (Constantinescu et al., 2016; Garvert et al., 2017), abstraction and transfer (Cole, Etzel, Zacks, Schneider, & Braver, 2011), reward predictions (Takahashi, Stalnaker, Roesch, & Schoenbaum, 2017), associative inference and and schema learning (Hebscher & Gilboa, 2016; McKenzie et al., 2014; Moscovitch & Melo, 1997; Spalding et al., 2018; van Kesteren et al., 2013; Yu, 2018; Zeithamova, Dominick, & Preston, 2012; Zeithamova & Preston, 2010). Previous work has proposed a hierarchy of time-scales in the brain (Chen, Hasson, & Honey, 2015) and indicated a role for hippocampal-prefrontal interactions in integrating episodes to build abstract schema (Schlichting & Preston, 2017). In previous modeling work, we have proposed a role for hierarchies of predictive representations along prefrontal and hippocampal gradients (Momennejad & Howard, 2018). Our findings are consistent with and advance these previous findings, suggesting that representations of relational structures–many of them graph-based–may be systematically distributed at various scales of abstraction–or

clustering–along PFC and the medial temporal lobe hierarchies. This evidence is consistent with the proposal that these predictive representations support cognitive maps, and provide the bedrock for more quantitative and detailed theories of cognitive maps and herald the revision of earlier perspectives (O'Keefe & Nadel, 1978; Tolman, 1948).

It is worth noting that the fMRI dataset used in the present analyses (Brunec, Bellana et al., 2018) was measured as participants moved through a virtual Google navigation of a city they lived in (Toronto). Importantly, here participants navigated realistically long spatial distances, between 1-5 kilometers, which allowed us to truly distinguish between different predictive scales. However, the data set has some caveats, some of which we addressed in our controlled analyses and some of which remain to be addressed by future studies.

One caveat of the present dataset was that the familiar routes were, on average, significantly longer compared to the GPS condition (Figure 2). To overcome this caveat, we first controlled for distance in one-sample t-tests to reveal regions with significant pattern similarity within a given horizon (Supplementary Figure 2). However, while these results confirmed earlier findings this control still did not exclude the large difference between the distances traveled in the two conditions. Therefore, we reran the analyses excluding longer routes and including only familiar routes that were within the range of distances in the GPS condition. A t-test between the PFC searchlight similarity results in the familiar and the GPS conditions revealed that while at lower predictive scales OFC showed the most difference, at longer predictive scales more dorsal and polar PFC regions displayed significantly different similarity between the two conditions as well (Figure 8). Taken together, these controlled analyses suggest that our main findings are reliable (compare to Table 1 and Figures 6 and 7), but future studies with a closer control of traversed distance are needed.

Note that here we only used correlation as a measure of pattern similarity. However, while this measure is prevalent in fmri pattern similarity analysis in the memory domain (Ezzyat & Davachi, 2014), other measures such as mutual information and mahalanobis distance have been shown to yield more robust findings (Walther et al., 2016). Future studies are required to test and validate these findings using other measures of similarity distance including mutual information.

In the present study the selection of routes did not include multiple past and future routes for each subgoal location, and multiple past routes for goal locations. Future studies with such a design would allow the further testing of the graph structure of relational structures. Such a study will also advance previous work using routes with multiple paths (Balaguer, Spiers, Hassabis, & Summerfield, 2016; Chanales, Oza, Favila, & Kuhl, 2017). Specifically, this design would allow us to dissociate pattern similarity due to the memory of the past from pattern similarity due to predictive representation, investigate graph-based clustering in the organization of relational knowledge, and enable testing whether graph-based compressions underlie abstraction in representational hierarchies and subgoal selection. Such a design can be easily

implemented in future fMRI studies, enabling a more thorough analysis of prefrontal-hippocampal interactions in abstraction, planning, and sub-goal processing.

Importantly, future studies that allow the construction of a larger graph of the state space have a number of potentials for advancing research. They would enable the analysis of the similarity between states given various paths between them, as well as searching for both policy-dependent and goal-independent representations of space. Previous studies had indicated the frontopolar cortex not merely in planning and prospective representations, but also in counterfactual reasoning (Boorman, Behrens, & Rushworth, 2011) and representing the value of alternative actions and plans (Boorman et al., 2009). It is therefore possible that the hippocampus and OFC rely on different scales of task-relevant state representations while the frontopolar cortex can track state-value relations in alternative and counterfactual courses of action as well. This is in line with the observation that lesions to the frontopolar cortex do not impair usual navigation or performance on intelligence or working memory tests, but severely impair the patient's ability for hierarchical task management, multi-tasking, and prospective memory in real life settings (Burgess, 2000; Volle, Gonen-Yaacovi, Costello, Gilbert, & Burgess, 2011) such as completing a sequential plan for simple everyday tasks, e.g., plan a visit to multiple stores on a street to write a note and stamp and post it (Burgess, 2000).

Another caveat is in distinguishing interpretations of our findings. Namely, it is possible that the observed similarity to successor states is due to to forward replay along previously replayed trajectories of cognitive maps (Ambrose, Pfeiffer, & Foster, 2016; Momennejad et al., 2018; Wu & Foster, 2014), the spread of activation across memory networks (Sievers & Momennejad, 2019), or an increased representational similarity between a given state and its frequently visited successor resulting in pattern similarity among cached representations (Ezzyat & Davachi, 2014; Garvert et al., 2017; Momennejad et al., 2017; Stachenfeld et al., 2017). The last case has also been discussed in other terms, such as increased association, integration, abstraction, clustering (Ritvo, Turk-Browne, & Norman, 2019). While there are clever analytic designs to hint one way or another, the temporal resolution of fMRI does not allow for a clear cut dissociation of these hypotheses. Future work using higher temporal resolutions such as electrophysiology, MEG, and other methods is required to test these distinctions.

A number of future directions can investigate the robustness of the present findings and theoretical proposal. Future studies can investigate subgoal processing in hierarchical planning by asking participants to navigate the same graph from different locations with goals at different scales (e.g., NY to California, Manhattan to Brooklyn, home to bodega at the corner). It is then possible to test the temporal hierarchy of large-scale predictive representations in the PFC for higher level plans and smaller subgoal processing in prepolar PFC regions and more detailed plans in the hippocampus. Whole-brain evidence points to separable brain networks corresponding to different spatial scales (Peer, Ron, Monsa, & Arzy, 2019). Notably, here we also investigated the similarity of the mean activation of a number of TRs in the beginning of the experiment to the goal state, and found higher similarity to goal in the familiar condition in both prefrontal and hippocampal regions (Supplementary Figure 1). This finding is in line with and

complements previous electrophysiology and neuroimaging work sowing goal representation in the hippocampus (Brown et al., 2016; Howard et al., 2014; Sarel, Finkelstein, Las, & Ulanovsky, 2017; Tsitsiklis et al., 2019). Future study designs with larger graphs–with multiple paths leading from one location to others–will be better suited to investigate the specifics of similarity to goal findings along the hippocampal-prefrontal hierarchies. Finally, follow up studies can more closely investigate abstraction to schema (McKenzie et al., 2014; Robin & Moscovitch, 2017; Schlichting & Preston, 2017; van Kesteren et al., 2013) that are formed due to PFC-medial temporal lobe interactions, and schema retrieval from PFC during hypothesis testing to bias hippocampal retrieval, e.g., entering a novel room, determining if it is a kitchen, office, or dining room, etc.

Summary We present support for the hypothesis that predictive maps with different scales are structured in hippocampal-prefrontal hierarchies. We found that while posterior hippocampal regions supported predictive scales–up to 100-200 meters–anterior prefrontal regions supported larger predictive horizons–which in this spatial navigation task extended to 875-900 meters. Our results support the idea that medial temporal-prefrontal interactions underlie cognitive maps and hierarchical planning. The organizational principles of predictive hierarchies can be extended to non-spatial domains (such as category and schema learning and abstraction). Follow up future studies can be specifically designed to investigate planning, subgoal setting, and the learning and use of schema in spatial and non-spatial settings.

References

Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. (2016). Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward. *Neuron*, *91*(5), 1124–1136. https://doi.org/10.1016/j.neuron.2016.07.047

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, *19*(12), 2082–2099. https://doi.org/10.1162/jocn.2007.19.12.2082

Balaguer, J., Spiers, H., Hassabis, D., & Summerfield, C. (2016). Neural Mechanisms of Hierarchical Planning in a Virtual Subway Network. *Neuron*, *90*(4), 893–903. https://doi.org/10.1016/j.neuron.2016.03.037

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415), eaat6766. https://doi.org/10.1126/science.aat6766

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, *62*(5), 733–743. https://doi.org/10.1016/j.neuron.2009.05.014

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual Choice and Learning

in a Neural Network Centered on Human Lateral Frontopolar Cortex. *PLOS Biology*, *9*(6), e1001093. https://doi.org/10.1371/journal.pbio.1001093

Brown, T. I., Carr, V. A., LaRocque, K. F., Favila, S. E., Gordon, A. M., Bowles, B., … Wagner, A. D. (2016). Prospective representation of navigational goals in the human hippocampus. *Science (New York, N.Y.)*, *352*(6291), 1323–1326. https://doi.org/10.1126/science.aaf0784

Brun, V. H., Solstad, T., Kjelstrup, K. B., Fyhn, M., Witter, M. P., Moser, E. I., & Moser, M.-B. (2008). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*, *18*(12), 1200–1212. https://doi.org/10.1002/hipo.20504

Brunec, I. K., Bellana, B., Ozubko, J. D., Man, V., Robin, J., Liu, Z.-X., … Moscovitch, M. (2018). Multiple Scales of Representation along the Hippocampal Anteroposterior Axis in Humans. *Current Biology*, *28*(13), 2129-2135.e6. https://doi.org/10.1016/j.cub.2018.05.016

Bunge, S. A., Kahn, I., Wallis, J. D., Miller, E. K., & Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *Journal of Neurophysiology*, *90*(5), 3419–3428. https://doi.org/10.1152/jn.00910.2002

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, *35*(4), 625–641. https://doi.org/10.1016/s0896-6273(02)00830-9

Burgess, P. W. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research*, *63*(3–4), 279–288.

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

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology: CB*, *27*(15), 2307-2317.e5. https://doi.org/10.1016/j.cub.2017.06.057

Chen, J., Hasson, U., & Honey, C. J. (2015). Processing Timescales as an Organizing Principle for Primate Cortex. *Neuron*, *88*(2), 244–246. https://doi.org/10.1016/j.neuron.2015.10.010

Christoff, K., & Gabrieli, J. D. E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*(2), 168–186.

Christoff, K, Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*(5), 1136–1149. https://doi.org/10.1006/nimg.2001.0922

Christoff, Kalina, Keramatian, K., Gordon, A. M., Smith, R., & Mädler, B. (2009). Prefrontal organization of cognitive control according to levels of abstraction. *Brain Research*, *1286*, 94–105. https://doi.org/10.1016/j.brainres.2009.05.096

Cole, M. W., Etzel, J. A., Zacks, J. M., Schneider, W., & Braver, T. S. (2011). Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. *Frontiers in Human Neuroscience*, *5*, 142. https://doi.org/10.3389/fnhum.2011.00142

Collin, S. H. P., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, *18*(11), 1562–1564. https://doi.org/10.1038/nn.4138

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science (New York, N.Y.)*, *352*(6292), 1464–1468. https://doi.org/10.1126/science.aaf0941

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*(12), 853–866. https://doi.org/10.1002/hipo.22993

Cox, R. W. (1996). RW Cox. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.

Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1655). https://doi.org/10.1098/rstb.2013.0478

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Duncan, K. D., & Schlichting, M. L. (2018). Hippocampal representations as a function of time, subregion, and brain state. *Neurobiology of Learning and Memory*, *153*, 40–56.

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*(11), 1504–1513. https://doi.org/10.1038/nn.4656

Evensmoen, H. R., Lehn, H., Xu, J., Witter, M. P., Nadel, L., & Håberg, A. K. (2013). The Anterior Hippocampus Supports a Coarse, Global Environmental Representation and the Posterior Hippocampus Supports Fine-grained, Local Environmental Representations. *Journal of Cognitive Neuroscience*, *25*(11), 1908–1925. https://doi.org/10.1162/jocn_a_00436

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*(5), 1179–1189. https://doi.org/10.1016/j.neuron.2014.01.042

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *ELife*, *6*. https://doi.org/10.7554/eLife.17086

Gilbert, S. (2011). *Decoding the content of delayed intentions. 31*, 2888–2894.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews. Neuroscience*, *7*(7), 523–534. https://doi.org/10.1038/nrn1931

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. https://doi.org/10.1016/j.neuropsychologia.2016.05.003

Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., … Spiers, H. J. (2014). The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. *Current Biology: CB*, *24*(12), 1331–1340. https://doi.org/10.1016/j.cub.2014.05.001

Javadi, A.-H., Emo, B., Howard, L., Zisch, F., Yu, Y., Knight, R., … Spiers, H. (2017). Hippocampal and prefrontal processing of network topology to simulate the future. *Nature Communications*, *8*, 14652. https://doi.org/10.1038/ncomms14652

Jung, M., Wiener, S., & McNaughton, B. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience*, *14*(12), 7347–7356. https://doi.org/10.1523/JNEUROSCI.14-12-07347.1994

Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J.-D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, *107*(13), 6010–6015. https://doi.org/10.1073/pnas.0912838107

Keiflin, R., Reese, R. M., Woods, C. A., & Janak, P. H. (2013). The Orbitofrontal Cortex as Part of a Hierarchical Neural System Mediating Choice between Two Good Options. *Journal of Neuroscience*, *33*(40), 15989–15998. https://doi.org/10.1523/JNEUROSCI.0026-13.2013

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., … Moser, M.-B. (2008). Finite Scale of Spatial Representation in the Hippocampus. *Science*, *321*(5885), 140–143. https://doi.org/10.1126/science.1157086

Koechlin, E. (2011). Frontal pole function: What is specifically human? *Trends in Cognitive Sciences*, *15*(6), 241; author reply 243. https://doi.org/10.1016/j.tics.2011.04.005

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science (New York, N.Y.)*, *318*(5850), 594–598. https://doi.org/10.1126/science.1142995

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science (New York, N.Y.)*, *302*(5648), 1181–1185. https://doi.org/10.1126/science.1088545

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: Evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, *72*, 341–372. https://doi.org/10.1016/j.pneurobio.2004.03.006

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science (New York, N.Y.)*, *315*(5814), 961–966. https://doi.org/10.1126/science.1135801

Lohnas, L. J., Duncan, K., Doyle, W. K., Thesen, T., Devinsky, O., & Davachi, L. (2018).

Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus. *Proceedings of the National Academy of Sciences*, *115*(31), E7418–E7427. https://doi.org/10.1073/pnas.1717088115

McDannald, M. A., Jones, J. L., Takahashi, Y. K., & Schoenbaum, G. (2014). Learning theory: A driving force in understanding orbitofrontal function. *Neurobiology of Learning and Memory*, *108*, 22–27. https://doi.org/10.1016/j.nlm.2013.06.003

McDannald, M. A., Takahashi, Y. K., Lopatina, N., Pietras, B. W., Jones, J. L., & Schoenbaum, G. (2012). Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. *The European Journal of Neuroscience*, *35*(7), 991–996. https://doi.org/10.1111/j.1460-9568.2011.07982.x

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal Representation of Related and Opposing Memories Develop within Distinct, Hierarchically Organized Neural Schemas. *Neuron*, *83*(1), 202–215. https://doi.org/10.1016/j.neuron.2014.05.019

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*, *20*(9), 1269–1276. https://doi.org/10.1038/nn.4613

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680–692. https://doi.org/10.1038/s41562-017-0180-8

Momennejad, Ida, & Haynes, J.-D. (2012). Human anterior prefrontal cortex encodes the “what” and “when” of future intentions. *NeuroImage*, *61*(1), 139–148. https://doi.org/10.1016/j.neuroimage.2012.02.079

Momennejad, Ida, & Haynes, J.-D. (2013). Encoding of prospective tasks in the human

prefrontal cortex under varying task loads. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(44), 17342–17349. https://doi.org/10.1523/JNEUROSCI.0492-13.2013

Momennejad, Ida, & Howard, M. W. (2018). Predicting the Future with Multi-scale Successor Representations. *BioRxiv*, 449470. https://doi.org/10.1101/449470

Momennejad, Ida, Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *ELife*, *7*, e32548. https://doi.org/10.7554/eLife.32548

Moscovitch, M., & Melo, B. (1997). Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia*, *35*(7), 1017–1034. https://doi.org/10.1016/s0028-3932(97)00028-6

Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences*, *112*(35), 11078–11083. https://doi.org/10.1073/pnas.1507104112

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Retrieved from https://repository.arizona.edu/handle/10150/620894

Pauli, W. M., Gentile, G., Collette, S., Tyszka, J. M., & O'Doherty, J. P. (2019). Evidence for model-based encoding of Pavlovian contingencies in the human brain. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-08922-7

Peer, M., Ron, Y., Monsa, R., & Arzy, S. (2019). Processing of different spatial scales in the human brain. *ELife*, *8*.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: The analysis of functional brain images*. Elsevier.

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, *17*(5), 230–240. https://doi.org/10.1016/j.tics.2013.03.005

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*, *59*(3), 2142–2154.

R Core Team. (n.d.). R: A language and environment for statistical computing (Version 3.5.1). Retrieved from https://www.R-project.org/

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews. Neuroscience*, *5*(3), 184–194. https://doi.org/10.1038/nrn1343

Ribas-Fernandes, J. J. F., Shahnazian, D., Holroyd, C. B., & Botvinick, M. M. (2018). Subgoal- and Goal-related Reward Prediction Errors in Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, 1–16. https://doi.org/10.1162/jocn_a_01341

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends in Cognitive Sciences*, *23*(9), 726–742. https://doi.org/10.1016/j.tics.2019.06.007

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: Hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences*, *17*, 114–123. https://doi.org/10.1016/j.cobeha.2017.07.016

Ruediger, S., Spirig, D., Donato, F., & Caroni, P. (2012). Goal-oriented searching mediated by ventral hippocampus early in trial-and-error learning. *Nature Neuroscience*, *15*(11), 1563–1571. https://doi.org/10.1038/nn.3224

Russek, E., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2018). FMRI

evidence for the successor representation in human value computation. *Society for Neuroscience*, 360.02. San Diego.

Sarel, A., Finkelstein, A., Las, L., & Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science*, *355*(6321), 176–180. https://doi.org/10.1126/science.aak9589

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Phil. Trans. R. Soc. B*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3–8. https://doi.org/10.1002/hipo.22523

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. https://doi.org/10.1038/ncomms9151

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. https://doi.org/10.1016/j.cobeha.2014.07.005

Schlichting, M. L., & Preston, A. R. (2017). The Hippocampus and Memory Integration: Building Knowledge to Navigate Future Decisions. In D. E. Hannula & M. C. Duff (Eds.), *The*

Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition (pp. 405–437). https://doi.org/10.1007/978-3-319-50406-3_13

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Sievers, B., & Momennejad, I. (2019). SAMPL: The Spreading Activation and Memory PLasticity Model. *BioRxiv*, 778563. https://doi.org/10.1101/778563

Simon, D. A., & Daw, N. D. (2011). Neural Correlates of Forward Planning in a Spatial Decision Task in Humans. *The Journal of Neuroscience*, *31*(14), 5526–5539. https://doi.org/10.1523/JNEUROSCI.4647-10.2011

Spalding, K. N., Schlichting, M. L., Zeithamova, D., Preston, A. R., Tranel, D., Duff, M. C., & Warren, D. E. (2018). Ventromedial Prefrontal Cortex Is Necessary for Normal Associative Inference and Memory Integration. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(15), 3767–3775. https://doi.org/10.1523/JNEUROSCI.2501-17.2018

Spiers, H. J., & Gilbert, S. J. (2015). Solving the detour problem in navigation: A model of prefrontal and hippocampal interactions. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00125

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653. https://doi.org/10.1038/nn.4650

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews. Neuroscience*, *15*(10), 655–669.

https://doi.org/10.1038/nrn3785

Takahashi, Y. K., Stalnaker, T. A., Roesch, M. R., & Schoenbaum, G. (2017). Effects of inference on dopaminergic prediction errors depend on orbitofrontal processing. *Behavioral Neuroscience*, *131*(2), 127–134. https://doi.org/10.1037/bne0000192

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Tsitsiklis, M., Miller, J., Qasim, S. E., Inman, C. S., Gross, R. E., Willie, J. T., … Jacobs, J. (2019). Single-neuron representations of spatial memory targets in humans. *BioRxiv*, 523753. https://doi.org/10.1101/523753

van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–2359. https://doi.org/10.1016/j.neuropsychologia.2013.05.027

Vikbladh, O. M., Meager, M. R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., … Daw, N. D. (2019). Hippocampal Contributions to Model-Based Planning and Spatial Memory. *Neuron*, *102*(3), 683-693.e4. https://doi.org/10.1016/j.neuron.2019.02.014

Volle, E., Gonen-Yaacovi, G., Costello, A. de L., Gilbert, S. J., & Burgess, P. W. (2011). The role of rostral prefrontal cortex in prospective memory: A voxel-based lesion study. *Neuropsychologia*, *49*(8), 2185–2198. https://doi.org/10.1016/j.neuropsychologia.2011.02.045

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods: Cognitive

maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience*, *advance online publication*. https://doi.org/10.1038/nrn.2016.56

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

Wimmer, G. E., & Büchel, C. (2019). Learning of distant state predictions by the orbitofrontal cortex in humans. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-10597-z

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews. Neuroscience*, *4*(2), 139–147. https://doi.org/10.1038/nrn1033

Wu, X., & Foster, D. J. (2014). Hippocampal Replay Captures the Unique Topological Structure of a Novel Environment. *The Journal of Neuroscience*, *34*(19), 6459–6469. https://doi.org/10.1523/JNEUROSCI.3414-13.2014

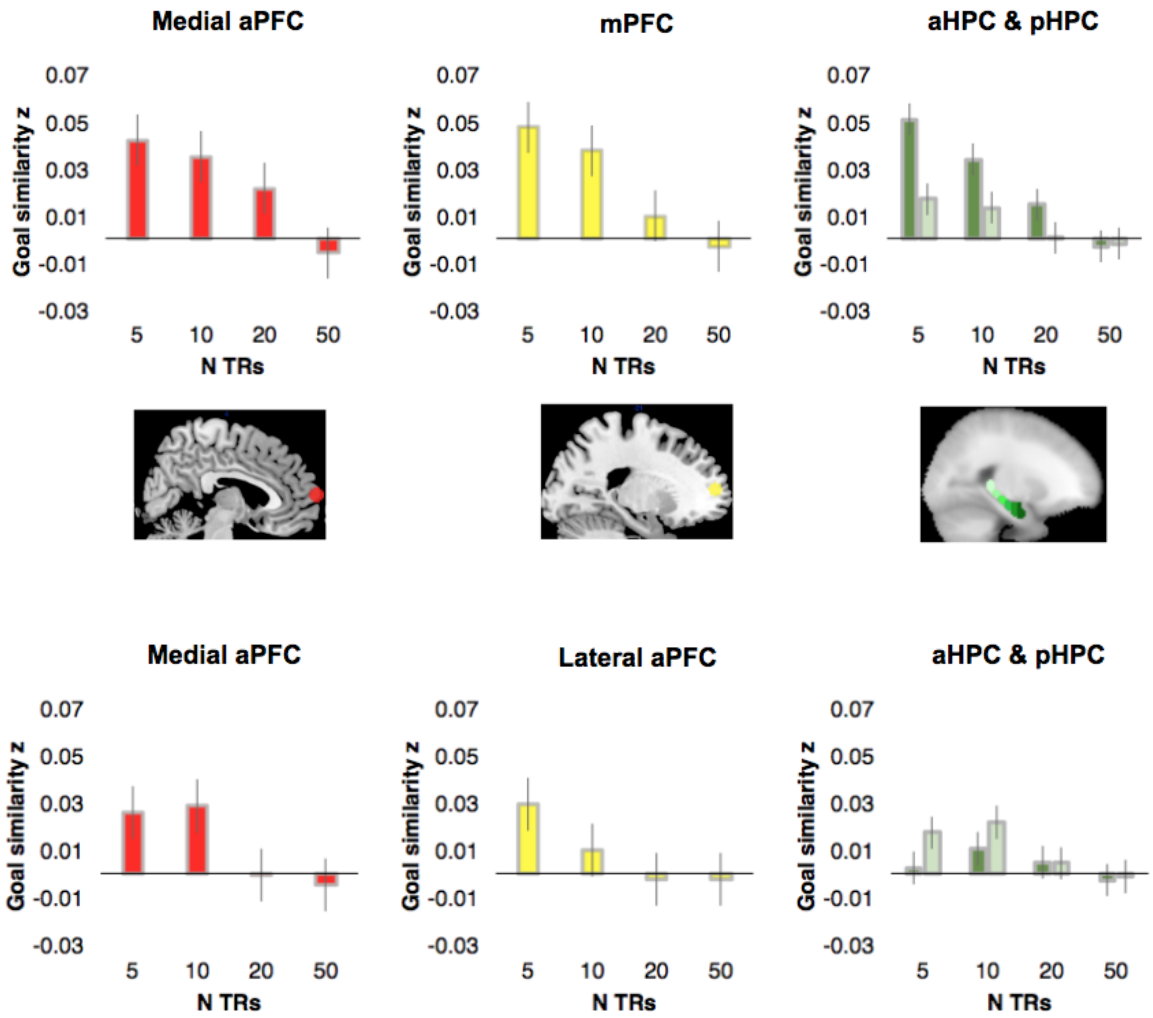
Yu, L. Q. (2018). Neural Mechanisms Underlying Schemas and Inferences. *Journal of Neuroscience*, *38*(37), 7930–7931. https://doi.org/10.1523/JNEUROSCI.1345-18.2018

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168–179. https://doi.org/10.1016/j.neuron.2012.05.010

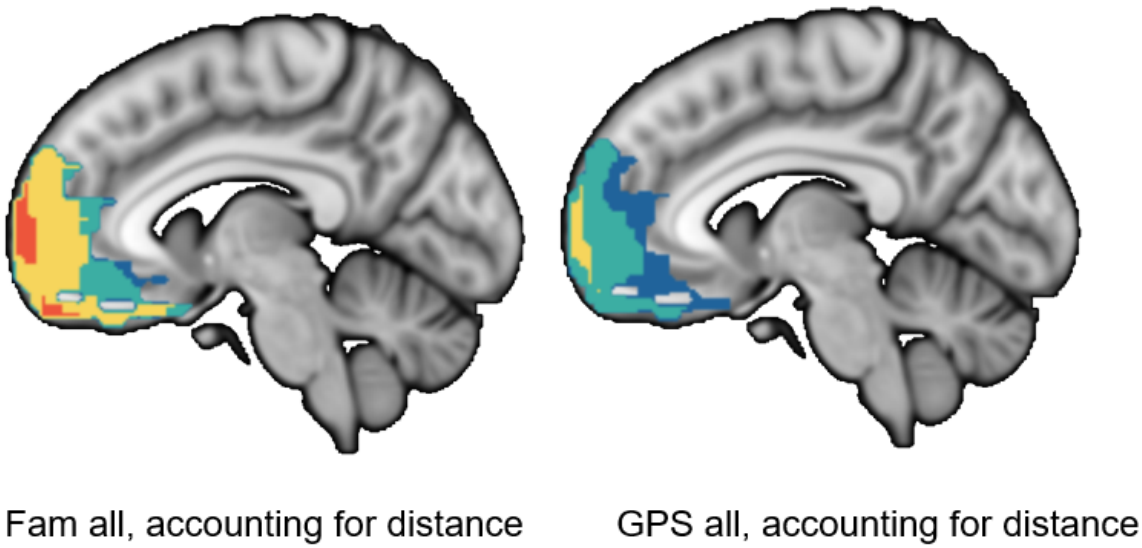
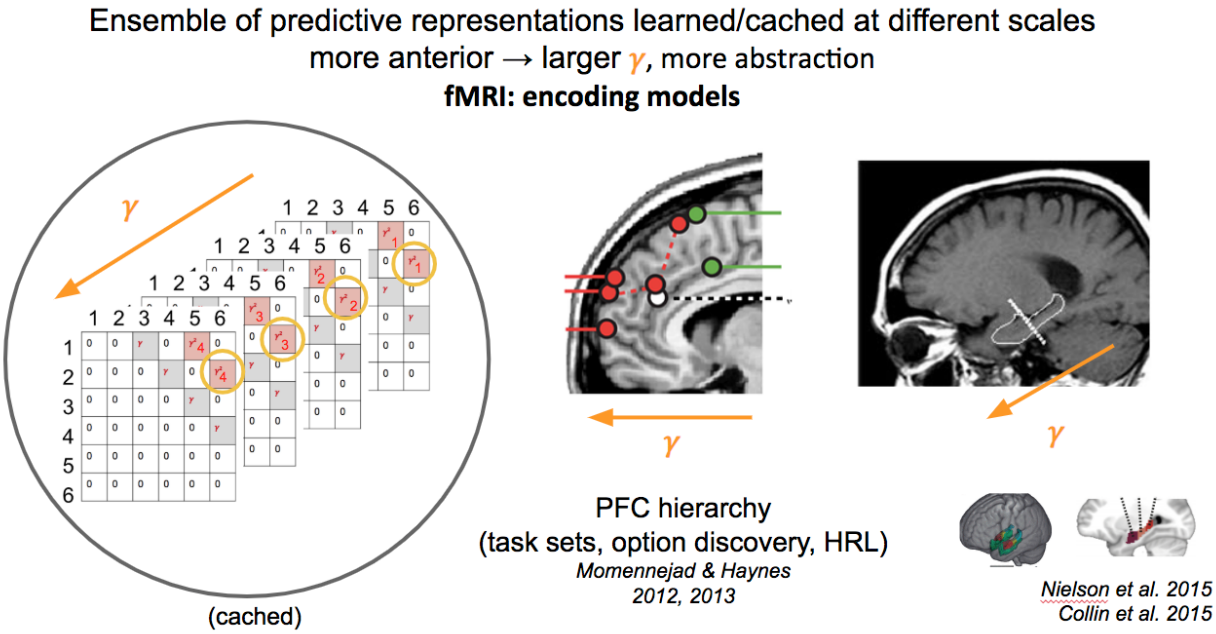
Zeithamova, D., & Preston, A. R. (2010). Flexible memories: Differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(44), 14676–14684. https://doi.org/10.1523/jneurosci.3250-10.2010

Supplementary Material

Supplementary Figure 1. Similarity of the fist n TRs to the goal state. We analyzed similarity of the first 5, 10, 20, and 50 TR to the goal state, and found significantly higher similarity to goal in the familiar conditions. In the familiar condition, PFC and anterior hippocampus contained more goal similarity up to 10 TRs. Only medial anterior PFC significantly sustained goal similarity within the first 20 TRs. In the GPS condition, small correlations of up to 10 first TRs was observed, which may be due to similarity of paths or goal search as participants nevigated an unknown path in a city they lived in and were familiar wiht. Red: anterior PFC, Yellow: mPFC, Dark green: Anterior hippocampus, Light green: Posterior hippocampus.



Supplementary Figure 2. One-sample T-tests with distance as covariate. The results look very similar to running to running a T-test on familiar routes vs. zero, and GPS routes vs. zero.[Comment from Iva: the clusters are smaller in the Fam condition, and the cluster for gamma=.8 is larger in GPS than the maps without distance covaried but generally the pattern is the same]



Supplementary Figure 3. Hypothesis schematic. Predictive representations with gradually increasing scale of abstraction long hippocampal and prefrontal hierarchies.