The brain hierarchically represents the past and future during multistep anticipation

Hannah Tarder-Stoll, Christopher Baldassano*, Mariam Aly*

Department of Psychology, Columbia University, New York, USA
*These authors contributed equally to this work

Please address correspondence to: Hannah Tarder-Stoll (hrt2116@columbia.edu)

ORCiD:
Hannah Tarder-Stoll: https://orcid.org/0009-0007-0957-4499
Christopher Baldassano: https://orcid.org/0000-0003-3540-5019
Mariam Aly: https://orcid.org/0000-0003-4033-6134

Acknowledgements:
This work was funded by a National Institutes of Health Research Project Grant (R01EY034436-01) and a Zuckerman Institute Seed Grant for MR Studies (CU-ZI-MR-S-0016) to M.A. and C.B. We would like to thank the Alyssano Group for helpful advice on this project.

Abstract

Memory for temporal structure enables both planning of future events and retrospection of past events. We investigated how the brain flexibly represents extended temporal sequences into the past and future during anticipation. Participants learned sequences of environments in immersive virtual reality. Pairs of sequences had the same environments in a different order, enabling context-specific learning. During fMRI, participants anticipated upcoming environments multiple steps into the future in a given sequence. Temporal structure was represented in the hippocampus and across visual regions (1) bidirectionally, with graded representations into the past and future and (2) hierarchically, with further events into the past and future represented in successively more anterior brain regions. Further, context-specific predictions were prioritized in the forward but not backward direction. Together, this work sheds light on how we flexibly represent sequential structure to enable planning over multiple timescales.

Introduction

Memory allows us to use past experience to generate expectations about the future. Integration of past information to predict future events enables efficient planning and flexible behavior in complex environments[1–4] and has been proposed to be a primary function of memory systems[5] and of the brain itself[6,7]. For predictions to usefully impact behavior, they should be represented on multiple timescales, allowing us to anticipate not just immediately upcoming events but also events further in the future. Furthermore, predictions that are relevant for the current context should be flexibly prioritized over those that are less relevant. For example, when riding the subway, it would be useful to anticipate multiple stations ahead on the relevant line, but we need not anticipate upcoming stops on other lines passing through the same stations. Such context-specific prediction may be supported by leveraging memories of past stops, which contextualize where we are in the present. Here, we aimed to test three central hypotheses. First, that the brain will flexibly anticipate events at multiple timescales in the future; second, that the future and the past will be represented simultaneously in the same brain regions; and third, that anticipatory representations will be prioritized for events that are contextually relevant.

To test these hypotheses, we drew on prior research showing anticipatory signals across the brain, particularly in memory and sensory systems[5,8–12]. For example, predictions about upcoming items or locations in a sequence are represented in visual cortex[13–16] and hippocampus[17–21], suggesting coordination between these regions in memory-based prediction of visual stimuli[16]. Although earlier research on prediction typically focused on one or a few brain regions[13,15–17,20] and predictions about immediately upcoming events[13,16,20], more recent work has shown that the brain represents anticipatory signals at multiple timescales simultaneously, with shorter timescales of prediction in more posterior regions and successively longer anticipatory timescales in progressively more anterior regions[12,22]. For example, during repeated viewing of a

movie clip, posterior regions like visual cortex primarily represent the current moment, while anterior regions such as the insula represent upcoming events multiple seconds into the future[12]. These findings of multistep anticipatory signals are generally consistent with computational theories that the brain builds models of the world that cache temporal information about successive events, with different predictive timescales in different brain regions (i.e., multi-scale successor representations[22,23]).

This research on multiscale anticipation in the brain complements earlier work showing hierarchical representations of *past* states[24]. Mirroring the predictive hierarchy for future states[12,22], information from the past lingers in the brain during ongoing experience, with shorter timescales of past information represented in posterior regions and longer timescales in anterior regions[24–28]. In the hippocampus specifically, temporal coding in the form of sequence reactivation extends both into the past and the future[29–32]. Furthermore, the brain's representations of the past and future can be flexibly modulated based on task demands[33]. Although this work suggests that the brain may represent both anticipated and past events, these prior studies did not test whether forward and backward representations of temporally extended structure existed simultaneously in the same brain regions. We therefore examined whether the brain contains bidirectional representations of the past and future, with the scale of these representations varying systematically across the brain.

For the brain's representations of temporal structure to be adaptive for behavior, they should flexibly change depending on context. Recent work in humans has shown context-specific patterns of activity in the hippocampus during goal-directed planning of future trajectories, suggesting that anticipation of temporally structured experience is specific to the upcoming items in a given context[34]. However, it remains unknown whether contextual modulation of temporal

structure representations is specific to planning trajectories in the forward direction or if contextual relevance also modulates representations of the past.

In the present study, we investigated how context-specific temporal structure is represented in the brain during a novel multistep anticipation task. Participants learned, in immersive virtual reality, four temporally extended sequences of eight environments each (**Figure 1**). Critically, pairs of sequences contained the same environments in a different order, requiring individuals to flexibly anticipate environments based on the current sequence context. Sequences were circular, such that environments were temporally predictable multiple steps into the future and the past regardless of location in the sequence. This allowed us to test whether temporal structure in both the prospective and retrospective direction is automatically represented in the brain even if only future states are task-relevant. Following sequence learning, participants were scanned with fMRI as they anticipated upcoming environments one to four steps into the future in a given (cued) sequence (Anticipation Task; **Figure 2**). Using multivoxel pattern similarity analyses in visual cortex, hippocampus, insula, and across the brain, we determined the extent to which temporal structure was (1) hierarchically represented along a posterior to anterior gradient, with further-reaching representations in more anterior regions; (2) represented in a bidirectional manner, with simultaneous representations of future and past environments within a context; and (3) modulated by context, with prioritized representations of nearby environments in the cued vs uncued sequence.

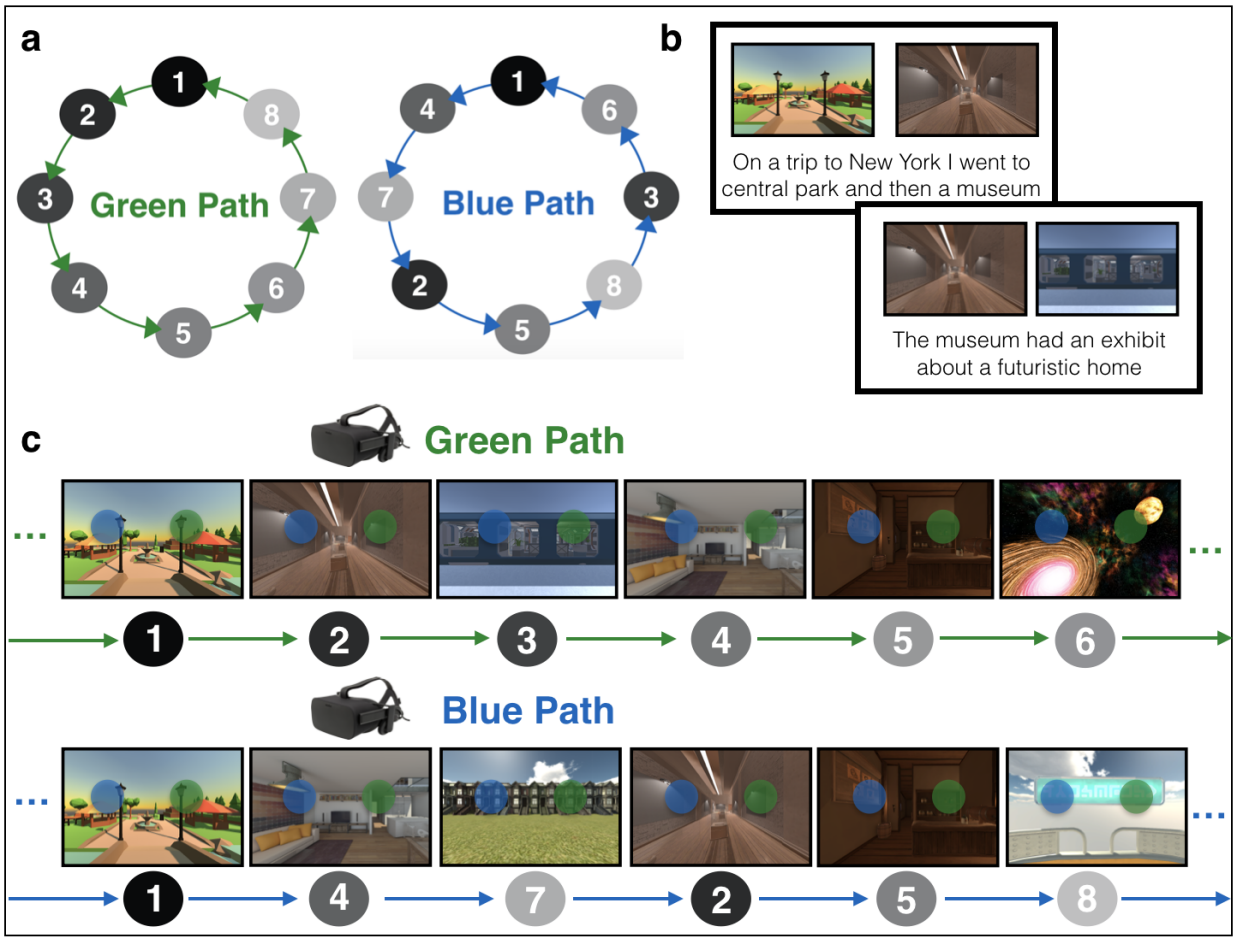

Figure 1 | Sequence structure and behavioral training. (a) Sequence structure. Participants learned sequences of eight environments, indicated by the gray nodes. The green path and the blue path consisted of the same environments in a different order. The sequences were constructed to be as distinct as possible: for a given environment the two preceding and two succeeding environments were different across the sequences. Participants learned four sequences in total: one green and blue path with a set of eight environments, and another green and blue path with a different set of eight environments. Only one green and blue path is depicted here for illustrative purposes. **(b)** Story Generation. To learn the sequence of environments, participants generated stories for each path to link the environments in order. Participants were told to link the final environment back to the first environment to create a loop. **(c)** Virtual Reality Training. Participants then explored the environments in immersive virtual reality in the green path order and the blue path order while rehearsing their stories. In a given environment, a green and blue sphere would appear. These spheres, when touched, teleported the participant to the next environment in the corresponding (green or blue) sequence. Participants then recalled the order of each of the four sequences (not shown).

Results

Anticipation Task Performance

Participants performed effectively on the Anticipation Task (**Figure 2a**), correctly choosing the closer of the two probe images, relative to the cued image and path, 86.86% of the time (sd = 0.08), which was significantly higher than chance performance of 50% ($t(31) = 61.08$, $p < 0.00001$). There was a trend toward higher accuracy in Map A (i.e. the first learned map) compared to Map B (beta = -0.283, 95% CI = [-0.576, 0.001], p = 0.058). Accuracy did not vary by path (Green vs Blue, beta = -0.124 , 95% CI = [-0.458, 0.209], p = 0.464), nor was there a map by path interaction (beta = 0.268 , 95% CI = [-0.274, 0.81], p = 0.333). Response times on the Anticipation Task were not significantly influenced by map (beta = 0.029, 95% CI = [-0.014, 0.072], p = 0.185), path (beta = 0.038, 95% CI = [-0.009, 0.085], p = 0.127), or their interaction (beta = -0.052, 95% CI = [-0.134, 0.029], p = 0.209).

We next determined how performance on the Anticipation Task varied by steps into the future (i.e. how many steps the correct probe was from the cue image on the cued path; also see[35]). Steps into the future had a trending impact on accuracy (beta = -0.15, 95% CI = [-0.33, 0.027], p = 0.096; **Figure 2b**), with an average difference in accuracy of 3.9% between one-step and four-step trials. Steps into the future robustly impacted response time (beta = 0.13, 95% CI = [-0.103 0.149], $p < 0.000001$; **Figure 2b**). Responses were on average 126 ms slower for each step into the future, with an average difference of 380 ms between one-step and four-step trials. Together, this suggests that participants performed accurately on the Anticipation Task, but were slower to anticipate upcoming environments that were further into the future.

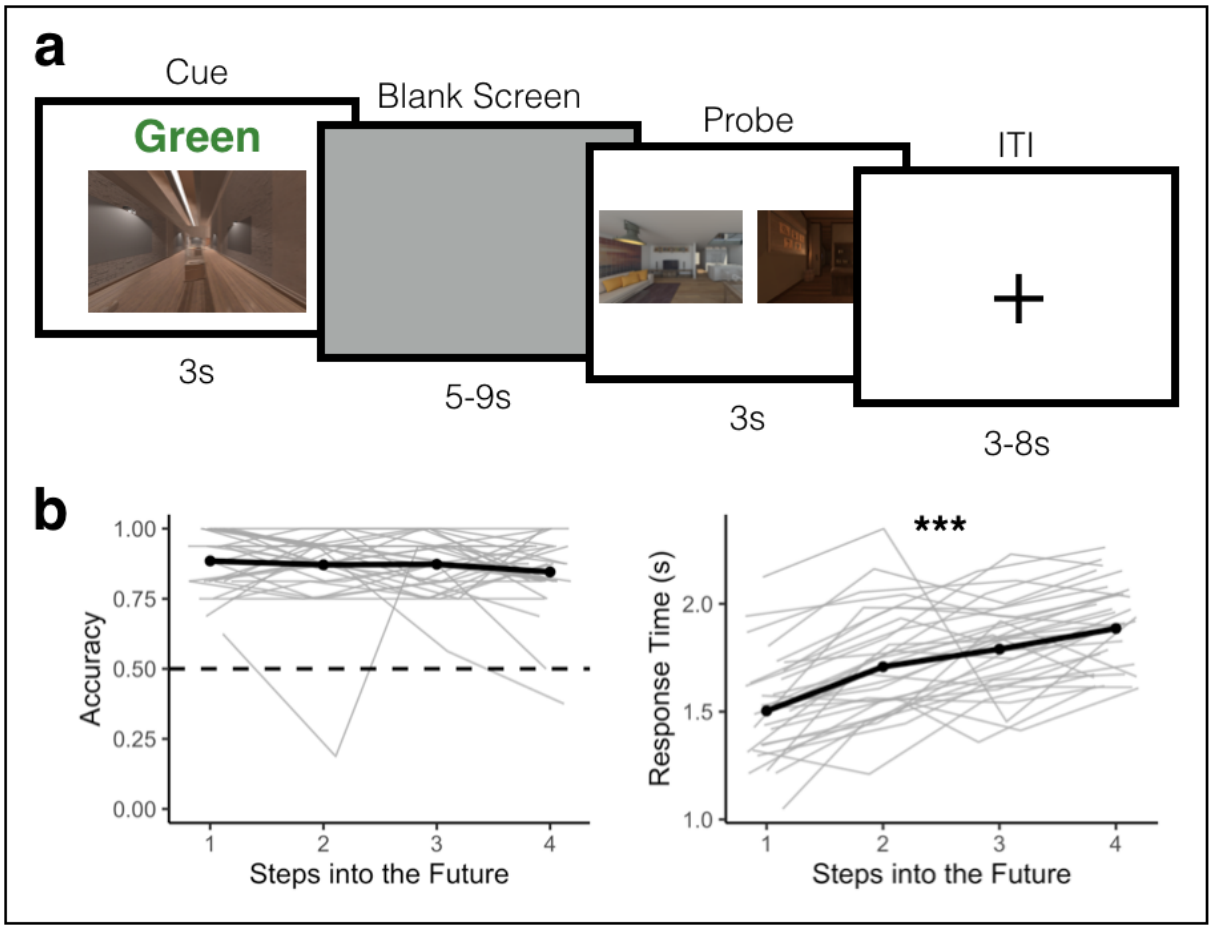

Figure 2 | Anticipation task and behavioral performance. (a) Anticipation Task. Participants returned one day after behavioral training and completed the anticipation task inside the MRI scanner. Participants were cued with a 2D image of an environment from one of the sequences along with a path cue (Green or Blue) for 3 seconds. They then saw a blank screen for a variable duration of 5 to 9 seconds during which they were told to anticipate upcoming environments. Participants were then probed with two images of upcoming environments and had 3 seconds to indicate which of the two environments was coming up sooner in the cued sequence, relative to the cue image. The correct answer could be 1 to 4 steps away from the cue image. **(b)** Behavioral Performance. Participants accurately anticipated upcoming environments in the cued sequence. Accuracy did not significantly differ across steps into the future (left). Response time, however, was significantly slower for further steps into the future (right). Pale gray lines indicate data for individual participants; the black line is the group mean. *** $p < .0001$

Bidirectional Representations of Temporal Structure in Hippocampus and Visual Cortex

For our MRI analyses, we first created conjunction ROIs by selecting voxels within visual cortex, hippocampus, and insula that reliably responded to distinct environments in the Localizer Task (**Figure 3a-c**). Next, we obtained the across-participant multivoxel pattern of brain activity for each environment within each of our ROIs (**Figure 3d**). To investigate neural representations of temporal structure during multi-step anticipation, we calculated pattern similarity between (1) multivoxel patterns of brain activity evoked during the Anticipation Task for each trial type (cue and path combination) for each participant and (2) the multivoxel patterns of brain activity evoked during Localizer Task, averaged across the remaining participants, for each environment on the same map (**Figure 4a**; see Methods). We then ordered the resulting correlation values in the sequence of the cued path with the cued environment in the center, successors following the cue to the right of the center, and predecessors to the left of the center (**Figure 4a**; see Methods). Importantly, because the order of environments in the sequences was randomized across participants, the across-participant multivoxel pattern of activity during the Localizer Task can not include information about successors for individual participants' sequences, resulting in a relatively pure measure of environment representations. Thus, this analysis allows us to determine the extent to which our regions of interest represented upcoming or preceding environments during the Anticipation Task, using activity pattern "templates" for each environment that were constructed to remove information about sequence structure.

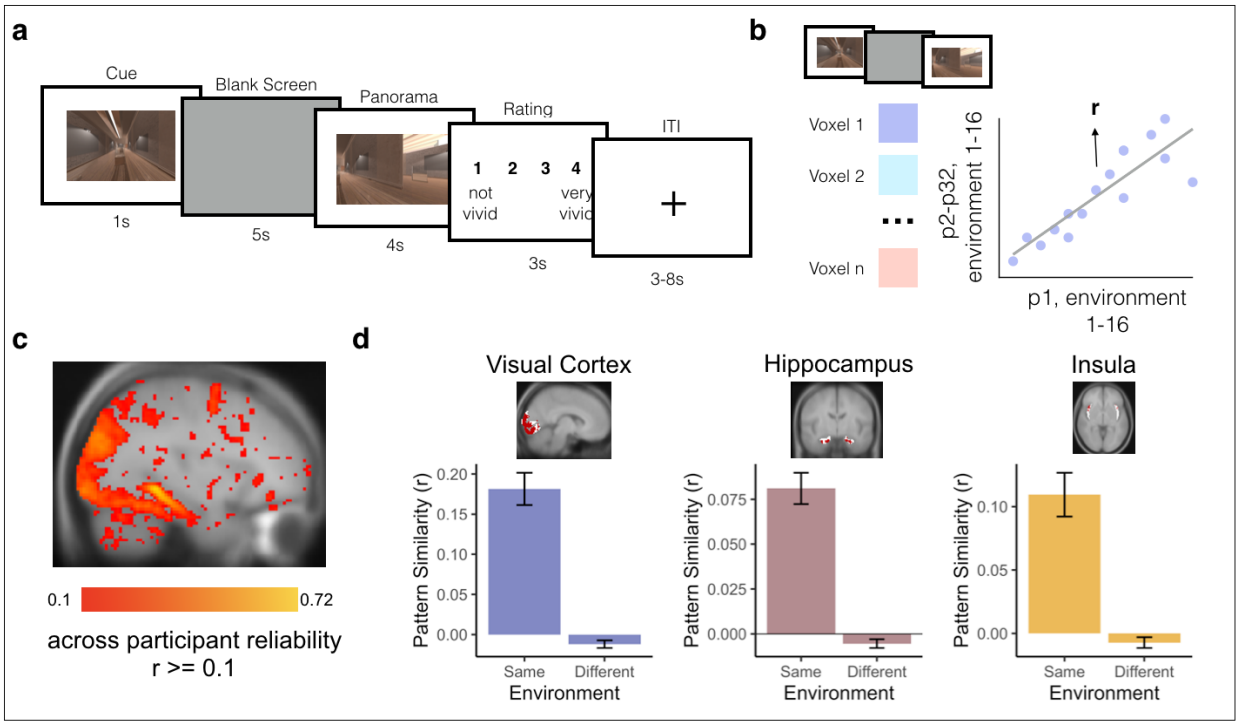

Figure 3 | Template brain activity patterns for each environment. (a) Localizer Task. Participants completed a localizer task inside the MRI scanner at the end of the session. Participants were cued with a 2D image of an environment from the experiment for 1 second. They then saw a blank screen for 5 seconds during which they were told to imagine being inside the environment in VR. Next, they saw images of the environment from different angles for 4 seconds and were given 3 seconds to rate how well their imagination matched the actual images of the environment. **(b)** Across-participant analysis for identifying voxels that reliably discriminate between environments. We measured the activity of each voxel in each participant during the Localizer Task (combining the cue, blank screen, and panorama phases) for each of the 16 environments. Next, we obtained the Pearson correlation (r) in each voxel between a participant's (e.g., P1's) responses to the 16 environments and the 16 average responses in the remaining participants (e.g., P2-P32). Averaging across all choices of the left-out participant, this yielded an across-participant reliability score for each voxel. **(c)** Whole brain map of voxels that reliably discriminate between environments. We only included voxels that had an environment reliability value of 0.1 or greater and were part of a cluster of at least 10 voxels in our subsequent analyses. **(d)** Environment reliability in ROIs. In visual cortex (left), hippocampus (middle), and insula (right), we selected the environment-reliable voxels (red) within each anatomically or functionally defined ROI (white). We then confirmed that the analysis successfully identified across-participant patterns of activity within these conjunction ROIs that were more correlated for the same environment than for different environments. Error bars indicate standard error of the mean.

We examined whether the sequence structure was reflected in the brain's representations during the Anticipation Task. We tested two main questions: first, whether these

representations were biased in the forward direction, suggesting stronger anticipation of future environments rather than representations of past environments; and, second, whether anticipatory representations were further reaching in some brain regions relative to others. To do so, we fit an asymmetrical Gaussian curve to the ordered pattern similarity values (see Methods). The Gaussian similarity model has four parameters: amplitude, asymptote, and forward and backward width (σ) (**Figure 4b**). The amplitude of the curve indicates the degree to which a brain region is representing the cue environment while it is on the screen. The forward and backward widths (σ) of the curve indicate how similarity to neighboring environments falls off with the number of steps in the forward and backward directions. Wider (vs. narrower) widths indicate that the brain region represents environments that are further away. If a brain region has a wide forward width but a narrow backward width, this indicates a bias towards representing upcoming environments, indicating anticipation, over retrospective representations of preceding environments. The asymptote is an indication of the representations of environments that are not captured by the width of the Gaussian; if the asymptote is lower than baseline (defined as pattern similarity between the cue and all environments from the other map, henceforth referred to as different-map baseline), this suggests that these environments are suppressed.

In visual cortex, the Gaussian model significantly outperformed the null models in which the order of the pattern similarity values was shuffled ($p < 0.001$). The amplitudes of participants' Gaussian fits were significantly higher than the different-map baseline, indicating that the cue environment was represented while it was on the screen (mean = 0.091, standard deviation = 0.029; $t(31) = 18.01$, $p < 0.000001$; **Figure 4c**). The asymptote was significantly lower than the different-map baseline, suggesting that other environments surrounding the cue were suppressed (mean = -0.014, standard deviation = 0.009; $t(31) = -5.97$, $p = 0.000001$; **Figure 4c**). The backward and forward widths (σ) were 0.712 steps and 0.634 steps, respectively, and were

not significantly different from each other (V(31)= 195.00, p = 0.203), suggesting that representations were not biased toward one direction over the other.

Turning to representations in hippocampus, the asymmetric Gaussian once again provided better fits than a permuted model (p = 0.029). In the hippocampus, similar to visual cortex, the amplitude of the Gaussian fit was significantly higher than the different-map baseline (mean = 0.0185, standard deviation = 0.030; t(31) = 2.959, p = 0.005; **Figure 4d**) and the asymptote was significantly lower than the different-map baseline (mean = -0.007, standard deviation = 0.014; t(31) = -3.476, p = 0.001; **Figure 4d).** Thus, hippocampus represented the cue environment while it was on the screen and suppressed environments further away. The backward and forward widths (σ) were 2.313 steps and 1.782 steps, respectively, and were again not significantly different from each other (V(31) = 231.00, p = 0.548), suggesting that the hippocampus represented forward and backward environments similarly. In contrast to hippocampus and visual cortex, Gaussian representations in insula were not significantly different from chance (p = 0.139).

We then statistically compared the width parameters in hippocampus and visual cortex, to determine whether hippocampus represented more distant environments surrounding the cue and whether this differed in the forward vs. backward direction. The width of the Gaussian in hippocampus was significantly wider than that in visual cortex (beta = 1.379, 95% CI = [0.559, 2.199], p = 0.0023; **Figure 4e**), indicating that hippocampus represented more distant environments. There was no effect of direction (beta = -0.246, 95% CI = [-0.935, 0.443], p = 0.489), nor a region by direction interaction (beta=-0.569, 95% CI = [-1.746, 0.606], p = 0.346. Together, this suggests that hippocampus had further reaching representations of temporal structure than visual cortex. However, representations were not biased toward the forward,

compared to the backward, direction in either region nor were there differential directional biases across regions.

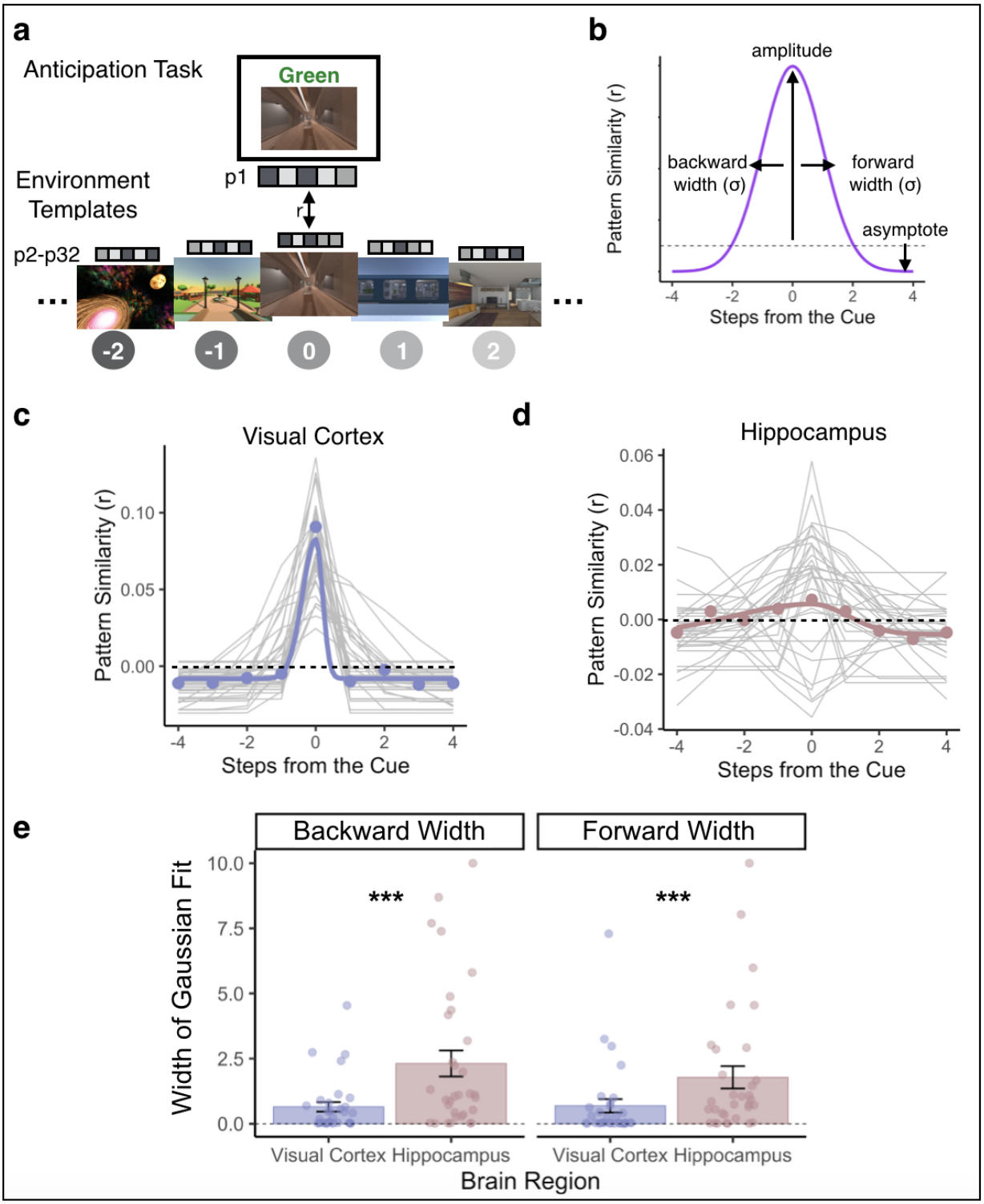

Figure 4 | Bidirectional and graded representations of temporal structure in hippocampus and visual cortex. (a) Schematic depiction of Gaussian analysis. We obtained the correlation between a given participant's (e.g., P1) cue screen activity pattern for each trial of the Anticipation Task and the remaining participants' (e.g., P2-P32) averaged patterns of activity for each of the

environment templates on the cued path. We then ordered the resulting pattern similarity values with the cue in the center and fit an asymmetrical Gaussian curve. **(b)** Gaussian similarity model. The amplitude of the curve is an indication of the degree to which a brain region is representing the cue environment while it is on the screen. The widths (σ) of the curve indicate how similarity to neighboring environments falls off with the number of steps in the forward and backward directions. Wider (vs. narrower) widths indicate that the brain region represents environments that are further away. The asymptote quantifies the representations of environments that are not captured by the width of the Gaussian; if the asymptote is lower than the dashed line (different-map baseline) this suggests that these environments are suppressed. **(c)** Gaussian curve in visual cortex. Visual cortex strongly represented the cue environment while it was on the screen (above-baseline amplitude) and did not strongly represent nearby environments (narrow forward and backward widths (σ)), instead showing suppression of environments other than the cue (below-baseline asymptote). Purple lines and points indicate the group-average pattern similarity values and Gaussian curve. Gray lines indicate each participant's Gaussian curve. **(d)** Gaussian curve in the hippocampus. The hippocampus represented the cue environment while it was on the screen (above-baseline amplitude), represented nearby environments in a graded manner, in both the forward and backward direction (wide forward and backward widths (σ)), and suppressed environments that were furthest away (below-baseline asymptote). Pink lines and points indicate the group-averaged pattern similarity values and Gaussian curve. Gray lines indicate each participant's Gaussian curve.* $p < .05$ **(e)** The width (σ) of the Gaussian curve in hippocampus was wider than that in visual cortex, indicating representations of environments further away. Widths did not significantly differ between the forward vs. backward directions in either visual cortex or the hippocampus. Bars indicate average width across participants, error bars indicate standard error of the mean, and small, transparent points indicate each participant's width estimates. *** $p < .001$

Hippocampal suppression of environment representations predicts response time costs

We next sought to examine whether neural representations of temporal structure were related to behavioral performance on the Anticipation Task. We reasoned that suppression of environments surrounding the cue, indicated by the asymptote parameters of the model fits, should interfere with the generation of long timescale predictions: more suppression should be associated with more response time costs for accessing future environments. To test this, we obtained the Spearman rank-order correlation between participants' Gaussian asymptote, separately for visual cortex and hippocampus, and the slope of their response times across steps

into the future. We hypothesized that a lower (more suppressed) asymptote would be related to steeper response time slopes across steps into the future, indicating a larger response time cost when making judgements about further environments. We expected this relationship to be stronger in hippocampus vs visual cortex, because visual cortex showed suppression of even the most nearby environments (**Figure 4c**).

In visual cortex, there was no relationship between the asymptote of the Gaussian curve and the response time slope across steps into the future (rho = -0.050, p = 0.784, **Figure 5a**). As hypothesized, in the hippocampus there was a significant negative correlation between the asymptote of the Gaussian curve and response time slope (rho = -0.362, p = 0.042, **Figure 5b)**, suggesting that suppression of environments surrounding the cue was related to response time costs for anticipating further environments.

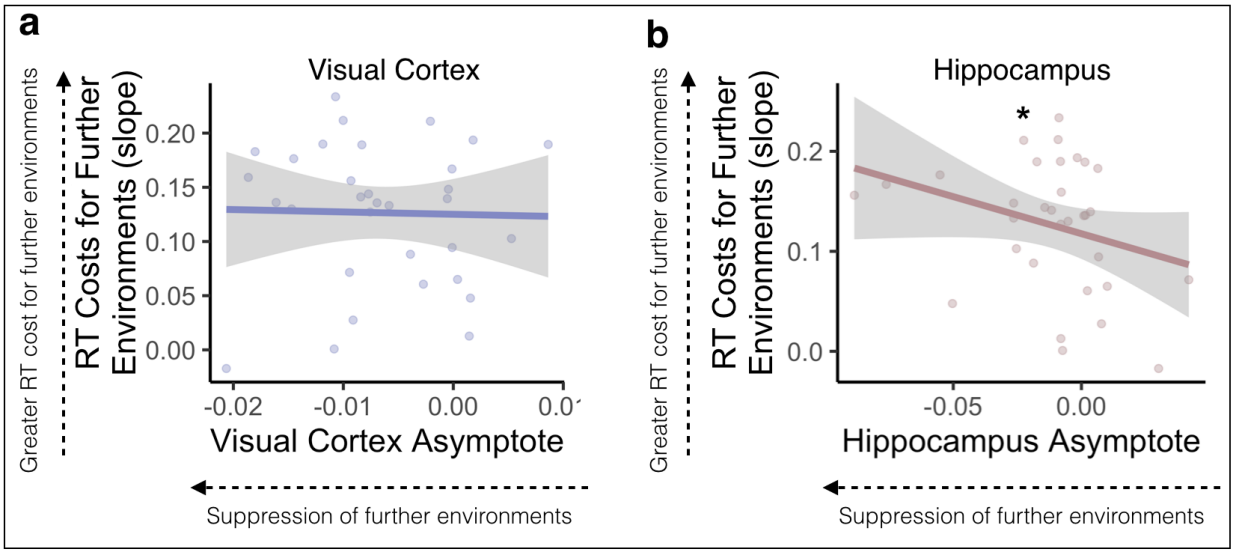

Figure 5 | Suppression of further environments in the hippocampus is related to response time costs. In visual cortex **(a)**, asymptotes, indicating suppression of non-cued environments, were not related to the slope of response times across steps into the future. In the hippocampus **(b)**, lower asymptotes were related to steep response time slopes, suggesting that participants were slower to respond to further environments when those environments were relatively suppressed. Lines and gray error ribbons indicate the correlation with 95% confidence intervals; points indicate each participant's asymptote and response time slope. * p < .05

Temporal structure is hierarchically organized within visual regions

We next conducted an exploratory searchlight analysis to determine which brain regions outside visual cortex and hippocampus exhibited Gaussian representations (see Methods). Our searchlight analysis revealed significant Gaussian representations across voxels in the visual system (**Figure 6a, Supplementary Figure 1**), including regions that code for scene information such as parahippocampal place area (PPA) and the retrosplenial cortex (RSC)[36]. There were no differences in backward vs forward widths (σ) in any voxel in the searchlight, suggesting bidirectional representations of temporal structure across the visual system.

Prior work has shown within-region functional differences in posterior vs anterior visual regions, including PPA and RSC[37,38]. Posterior aspects of these regions may play a larger role in scene perception while anterior aspects may represent scene memories. Based on these differences, we hypothesized that there may be hierarchical representations of temporal structure within PPA and RSC, with further reaching representations (as indicated by wider vs. narrower widths (σ)) in successively more anterior aspects of these regions. To test for hierarchical organization of temporal structure, we obtained the correlation for each participant between (1) the averaged forward and backward widths (σ) of the Gaussian curve in each voxel and (2) that voxel's y-coordinate, indicating its position along the posterior-anterior axis. We then tested whether these correlations were different from 0 across participants. We conducted the same analysis for the amplitude and asymptote, to determine if the representation of the cued environment and suppression of nearby environments also changed along the posterior-anterior axis.

There was a significant positive correlation between width (σ) and y-coordinate, indicating that Gaussian fits became progressively wider in progressively more anterior aspects, in both PPA ($t(31 = 2.424$, $p = 0.021$) and RSC ($t(31) = 2.638$, $p = 0.013$). There was a negative correlation

between amplitude and y-coordinate in PPA (t(31) = -2.636, p = 0.013), but not RSC (t(31)=-0.550, p=0.586). Finally, there was no correlation between asymptote and y-coordinate in either PPA (t(31) = 1.721, p = 0.095) or RSC (t(31) = 1.047, p = 0.303). This suggests a within-region hierarchical organization of representations in the visual system, such that more anterior (vs posterior) aspects of PPA and RSC represent environments that are further away in the past and future. Importantly, because amplitude was not correlated with y-coordinate in RSC, this suggests that further reaching representations are not necessarily a consequence of reduced processing of the present.

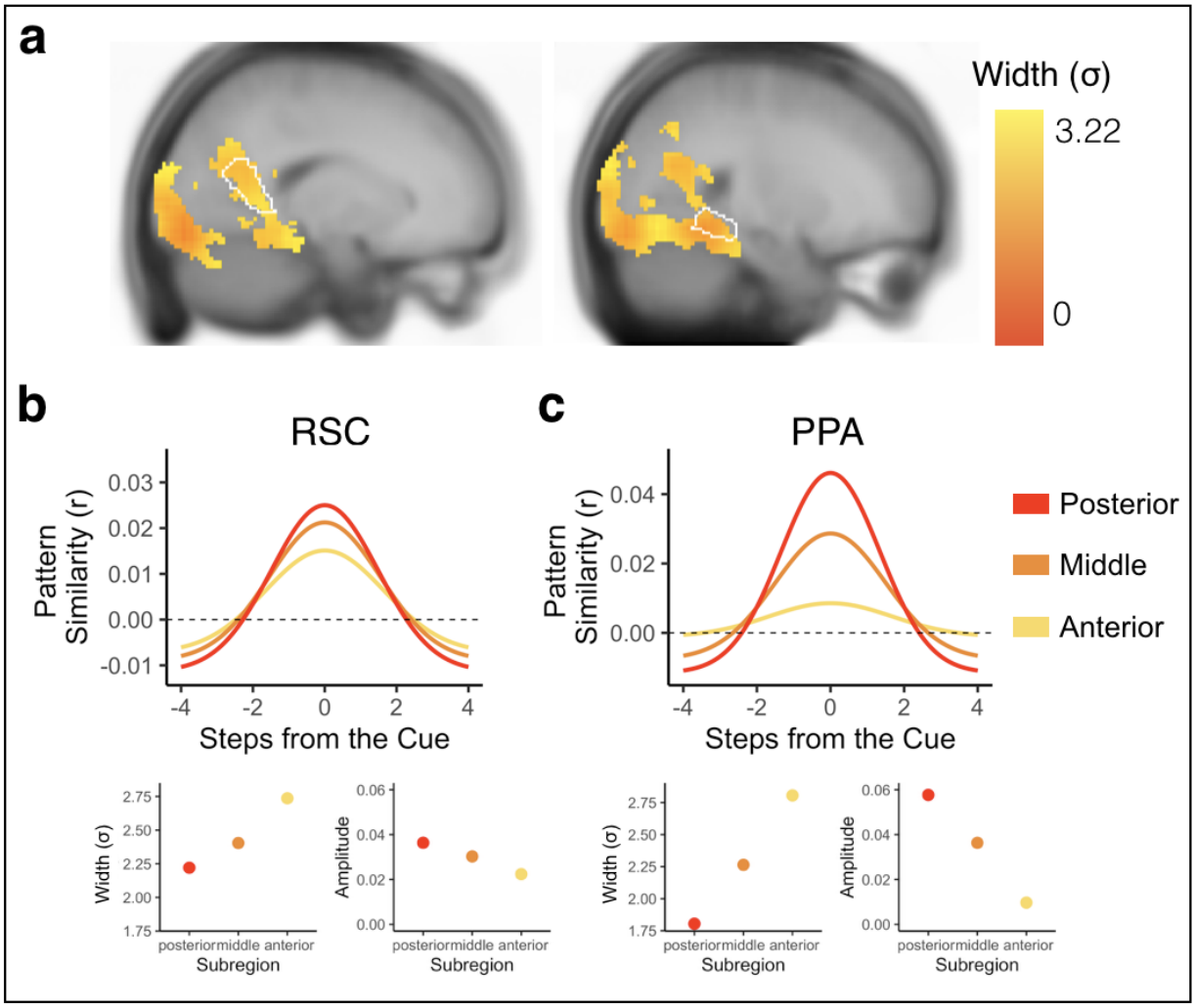

Figure 6 | Bidirectional and graded representations of temporal structure reveals within-region hierarchies in the visual system. (a) Searchlight results revealed statistically reliable Gaussian

representations in voxels across visual regions. Forward and backward widths (σ) of the Gaussian curves were hierarchically organized within visual regions (e.g., RSC and PPA), with narrow widths (indicated in red) in more posterior aspects of the region and progressively wider widths (indicated in yellow) in progressively more anterior aspects of the region. Gaussian fits of sample voxels are shown from RSC **(b)** and PPA **(c)**. Voxels in progressively more anterior (indicated in yellow) compared to posterior (indicated in red) aspects of RSC and PPA had progressively wider widths (bottom left in **(b)** and **(c)**) and progressively lower amplitudes in PPA (bottom right in **(c)**) but not RSC (bottom right in **(b)**).

Context-dependent representations of temporal structure

Having established how the brain represents temporal structure within a given sequence, we next sought to establish whether sequence representations are context dependent. We addressed this question by leveraging pattern similarity differences between the Green and the Blue path, which contained the same environments in a different order. We conducted pattern similarity analyses by obtaining the correlation between the multivoxel activity pattern evoked during a specific trial type (cue and path) for a given participant and the averaged multivoxel activity patterns from the Localizer Task from remaining participants for environments (1) one and two steps away, (2) in the forward and backward direction, and (3) on the cued and the uncued path (**Figure 7a**). We focused on one and two steps because our sequences were specifically designed so that environments one and two steps away, in both the forward and backward direction, were not shared between the two paths. We then obtained the difference in pattern similarity values between the cued and uncued path for each step and direction. We then tested whether the cued vs uncued difference in pattern similarity was influenced by step (one vs. two), direction (forward vs. backward), and their interaction in each of our three ROIs (visual cortex, hippocampus, and insula). Finally, we assessed whether the cued vs uncued difference in pattern similarity was different from zero, separately by trial type and ROI.

In the insula, there was a main effect of direction (beta = 0.033, 95% CI = [0.011 0.054], p = 0.003; **Figure 7b**), with higher cued vs uncued pattern similarity in the forward compared to the

backward direction. There was no main effect of step (beta = 0.016, 95% CI = [-0.005 0.038], p = 0.129), nor a step by direction interaction (beta = -0.010, 95% CI = [-0.052 0.032], p = 0.644). Given the main effect of direction, we conducted follow-up t-tests to assess whether context-dependent representations were significantly different from zero in the forward or backward directions. We found context-dependent prioritization of upcoming environments (t(31) = 2.443, p = 0.017), but not past environments (t(31) = -1.804, p = 0.076), showing a context-dependent forward bias in the insula.

In contrast, neither the hippocampus nor visual cortex showed modulation of context-dependent pattern similarity values by direction (hippocampus: beta = 0.004, 95% CI = [-0.011, 0.019], p = 0.589; visual cortex: beta = 0.009, 95% CI = [-0.002 0.021], p = 0.111; **Figure 7b**), step (hippocampus: beta = -0.011, 95% CI = [-0.024, 0.001], p = 0.095; visual cortex: beta = 1.235e-03, 95% CI = [-0.010, 0.013], p = 0.835), or their interaction (hippocampus: beta = -0.010, 95% CI = [-0.031, 0.011], p = 0.342; visual cortex: beta = 0.003, 95% CI = [-0.019 0.027], p = 0.745). Furthermore, the differences in pattern similarity values for the cued vs. uncued paths were not different from zero for either step in either direction for either region (all *p*s > 0.08), indicating that these regions did not show context-dependent representations during the cue screen.

Although we failed to find evidence for context-dependent representations in visual cortex during the cue screen of the Anticipation Task, this may have been because a visual stimulus was presented: visual cortex activity is strongly modulated by visual input. Thus, we tested whether visual cortex representations during the Blank Screen period of the Anticipation Task (**Figure 2a**) showed evidence for anticipation[14]. We focused on context-dependent anticipation of the immediately upcoming environment because of past work showing one-step anticipatory signals in visual cortex[15,16]. We obtained the correlation between (1) multivoxel patterns from the blank screen of the Anticipation task for each trial type (cue and path) for a

single participant and (2) the averaged multivoxel patterns for the remaining participants for the environment templates coming up one step in the forward direction on the cued and uncued paths (**Figure 7a**). We then obtained the difference in the cued and uncued pattern similarity values. As hypothesized, we found that visual cortex represented the upcoming environment one step into the future on the cued path more than the uncued path (t(31) = 2.079, p=0.046; **Figure 7c**). Thus, visual cortex represented one step context-dependent predictions, but only in the absence of visual stimulation.

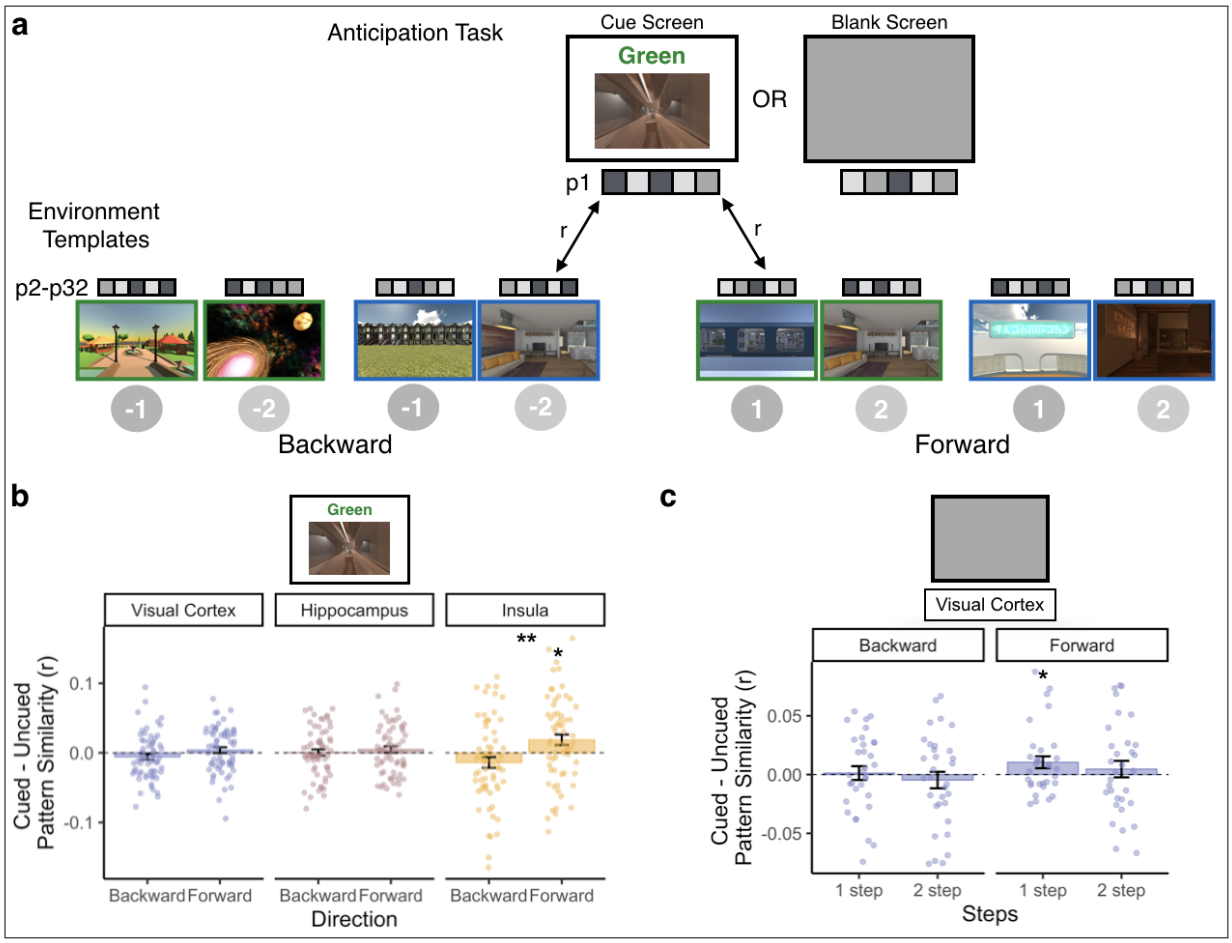

Figure 7 | Context-dependent representations of multistep anticipation. (a) Schematic depiction of context-dependence analysis. We obtained the correlation between a given participant's (e.g., P1) cue screen (or blank screen) activity patterns from the Anticipation Task and the remaining participants' (e.g., P2-P32) averaged patterns of activity for environment templates (1) one and two steps away, (2) in the forward and backward direction, and (3) on the cued and uncued path. We then subtracted the correlations on the cued and uncued path (e.g., one step in

the forward direction on the cued path minus one step in the forward direction on the uncued path) to obtain a measure of context-dependent pattern similarity. Positive values indicate prioritized representations for environments on the cued path, whereas negative values indicate suppressed representations for environments on the cued path. **(b)** Cue screen analysis. During the cue screen, the insula represented environments one and two steps in the forward direction in a context-dependent manner, with stronger representations for upcoming environments on the cued vs. the uncued path. Yellow bars indicate group-average pattern similarity values and yellow points indicate individual participants' pattern similarity values. Error bars indicate standard error of the mean. The hippocampus (indicated in pink) and visual cortex (in purple) did not represent nearby environments in either direction in a context dependent manner. Data are collapsed across steps for visualization purposes. **(c)** Blank screen analysis. During the blank screen, visual cortex more strongly represented one step in the forward direction on the cued vs. the uncued path. Visual cortex did not show context-dependent representations in the backward direction. Purple bars indicate group-average pattern similarity values and purple points indicate individual participants' pattern similarity values. Error bars indicate standard error of the mean.

Discussion

We examined how extended temporal structure is represented in the brain during context-dependent anticipation of future events. Participants anticipated multiple steps into the future accurately, but were slower to anticipate far vs near events. Multivoxel fMRI analyses revealed bidirectional representations of temporal structure within a context, with graded representations of environments in the forward and backward direction. Visual cortex (V1-4) primarily represented the current environment, whereas hippocampus represented further environments in a graded fashion. Hippocampal representations of temporal structure were relevant for behavior: suppression of distant environments was linked to response time costs for anticipating further events. Beyond visual cortex and hippocampus, a hierarchy of temporal structure was also apparent within individual visual regions: successively more anterior aspects of PPA and RSC represented further environments into the past and future. Finally, visual cortex and insula exhibited context-dependent representations in the forward but not backward direction, suggesting prioritization of context-relevant information about future states.

Our results are generally consistent with influential theories of prediction in the brain. Graded coding of upcoming events is consistent with successor representation models[2,23,39,40], which suggest that information about future states becomes cached into the representation of the current state in a temporally discounted manner. These models have been extended to account for multiple timescales of prediction by incorporating different scales of temporal discounting[23]. In line with these theories, recent work has shown that multiple timescales of prediction are represented simultaneously in the brain[12], with less evidence for further predictions[22,41]. Strikingly, although our asymmetric Gaussian analysis was designed to allow differential coding of the future vs the past, representations were not uniquely biased toward future states. Instead, the hippocampus and visual system represented temporal structure bidirectionally, with graded representations into the past and future. Taken together with prior work showing that hippocampal representations of temporal distance[41–43] can be flexibly biased in either the forward or backward direction based on task demands[33], our findings suggest that representations of the past and future can exist simultaneously within the hippocampus, even though the task demands were to anticipate future states. Thus, our work extends theories of prediction across the brain, suggesting that graded retrospection of past states can occur alongside prediction of future ones.

An important distinction between our experiment and past studies of prediction is that our sequences were circular and temporally extended, whereas sequences in prior studies tended to have a clear end point (i.e. were linear instead of circular)[12,22,34,41,44] or were shorter[19]. Our novel design therefore allowed us to detect temporally extended representations in both the forward and backward directions. Thus, it is possible that prior work emphasizing anticipatory coding in the brain missed simultaneous representations of the past because the tasks were not designed to detect bidirectional temporal coding.

In addition to representing nearby environments in the past and future, we also found that the hippocampus suppressed more distant environments, showing deactivation of these environments' patterns relative to an unrelated-environment baseline. Although suppressing distant environments can be beneficial for responding to imminent events, it can also lead to behavioral costs. Indeed, hippocampal suppression of distant environments was related to response times costs for anticipating further events. This highlights a trade-off between prioritizing nearby events and being able to quickly respond to upcoming events further in the future.

Representations of temporal structure extended beyond hippocampus and visual cortex. In an exploratory whole-brain searchlight analysis, we found representations of temporal structure across the visual system, including PPA and RSC, regions that play an important role in spatial cognition[36]. Both PPA and RSC represented the cued environment but also represented the temporal structure of surrounding environments in the sequence in both the forward and backward direction. Our findings therefore extend prior work showing that PPA responses can be modulated by temporal context[45] and prior contextual associations more generally[46–49]. Notably, our findings go beyond this prior work by showing a gradual progression of sequence coding within PPA and RSC, with progressively more anterior regions representing more of the future and past and less of the present. This is broadly consistent with prior work suggesting a posterior vs. anterior division within PPA, with posterior aspects playing a larger role in scene perception and anterior aspects playing a larger role in scene memory[37,38]. Thus, we show that, within a context, visual regions may balance representations of perception and memory, gradually incorporating less information from perception and more information about learned temporal structure along a posterior to anterior hierarchy.

To investigate bidirectional context-dependent sequence representations, we carefully manipulated overlapping sequences. Pairs of sequences contained the same environments in a different order, with no overlap in environments one and two steps into the future and the past. Prior work has found context-specific prediction of outcomes in visual cortex[16,50], which dovetails with our finding that visual cortex represented one-step into the future more strongly on the cued compared to the uncued sequence in the absence of visual stimulation. We also found that the insula represented two steps into the future in a context-dependent manner, furthering recent work from our lab showing long-timescale predictions in this region[12]. These findings are also consistent with theories that anterior brain regions should predict further into the future than posterior ones[22,23]. Interestingly, we only observed context-dependent prioritization into the future, but not the past. This suggests that prediction of future states may be context-specific, whereas retrospection may be context-independent. Another possibility is that context-specific prioritization of future or past states may depend on task demands: the demand to predict future states may have elicited context-dependent prediction but not retrospection. Future work could further examine the circumstances under which context-dependent prioritization of future and past states emerge.

Unlike visual cortex and insula, we were unable to detect context-specific representations in the hippocampus. This result is seemingly in contrast to past work showing context-dependent prediction in the hippocampus multiple steps into the future[34], and a large literature showing contextual representations across the hippocampus more generally (for reviews see[51,52]). One possibility for this discrepancy is that some hippocampal subregions emphasize integration across contexts whereas others emphasize differentiation[53–58]. We focused our analysis on voxels that show discriminable environment-specific activity patterns when no context cue is presented. These voxels may have been those that emphasize contextual integration in the hippocampus,

and may therefore automatically activate associated events whether or not they are relevant in the current context. Future work can separately examine anterior vs posterior hippocampus, or the CA3 vs dentate gyrus subfields, to determine whether context-dependent anticipation in the hippocampus is more likely in the latter vs former regions.

Broadly, it may be advantageous to represent temporal structure bidirectionally, rather than only prioritizing future states. For example, representing past states and future states could be a useful strategy when events surrounding ongoing experience differ based on context. Activating links toward past states as well as future ones may allow individuals to contextualize their current location within the sequence[24]. This possibility is consistent with our prior work showing that individuals represent sequences in terms of context-specific links between environments[35]: when an environment is cued, its associated links in both directions may be brought to mind so that the entire context is prioritized. An alternative possibility is that representing temporal structure into the past and future happens automatically: activating a particular moment within a temporally extended experience could cause activation to spread to the entire event representation[59–61], which may comprise both past and future. Future work could disentangle these possibilities and further investigate the circumstances under which future and past states are simultaneously represented.

Overall, the results presented here show that temporal structure is represented bidirectionally in the hippocampus and visual system. Future and past representations of temporal structure were graded, with less evidence for further environments in both the forward and backward direction, and were organized along a posterior to anterior hierarchy within and across regions. Our results further our understanding of how temporal structure is represented in the brain: such bidirectional representations could allow integration of past events from memory

alongside anticipation of future ones, which could support adaptive behavior during complex, temporally extended experiences.

Methods

Participants

Thirty-five healthy younger adults from the Columbia University community participated in the experiment. All participants gave written, informed consent in accordance with the Institutional Review Board at Columbia University. Participants were compensated $15 per hour for the behavioral training session and $20 per hour for the fMRI session (approximately $80 combined across both sessions). Three participants were excluded for technical issues with data collection, excessive motion (10% of TRs across all runs of the experiment marked as motion outliers by fMRIprep output), and dizziness inside the MRI scanner. Applying these exclusions resulted in a final sample of 32 participants (21 female, 19-35 years old, mean = 24.17, sd = 4.11, 13-29 years of education, mean = 16.85, sd = 3.76).

Overview

Participants learned two sequences ("Green Path" and "Blue Path") within each of two maps (Map A and Map B; **Figure 1a**). Map A and Map B contained eight distinct environments each. Within each map, the Green Path and Blue Path contained the same environments in a different order.

Participants first learned the order of the four sequences of environments by generating stories (**Figure 1b)** and then experiencing the environment sequences in immersive virtual reality using an Oculus Rift (**Figure 1c**). Participants returned 1 day later and completed the Anticipation Task in the MRI scanner (two runs, 32 trials per run). In the Anticipation Task, participants used

their memory for the four sequences to anticipate upcoming environments. They then completed a localizer task to obtain multivoxel patterns of brain activity for each environment (four runs, 16 trials per run).

Stimuli and Sequence Structure

Stimuli consisted of 16 3D virtual reality environments in the Unity game engine. Environments were obtained from asset collections in the Unity Asset Store. Half of the environments were indoor and half of the environments were outdoor. Using Unity, we created 2D images of each environment by rotating a virtual camera to eight different angles, 45 degrees apart. One angle was selected to be used as the cue and probe images throughout the task and the other angles were used for the panorama phase of the Localizer Task.

The 16 environments were used to form four sequences (Map A Green Path, Map A Blue Path, Map B Green Path, Map B Blue Path). Eight of the environments were assigned to Map A (i.e., the first learned set of environments) and the other eight environments were assigned to Map B (i.e., the second learned set of environments). Then, within each map (A or B), the Green Path and the Blue Path consisted of the same eight environments in a different order. The final environment in each sequence connected back to the first environment, forming a circle. The Green and Blue Paths were designed to be as distinct as possible: for a given environment the two preceding and two succeeding environments were different across the paths (**Figure 1a)**. The environment-to-map assignment and the order of the environments within a sequence was randomized across participants, although the Green Path was always shuffled in the same way to create the Blue Path, as described above.

Procedure

Participants first completed a training phase outside the MRI scanner. They returned one day later and completed a sequence refresher task outside the MRI scanner before taking part in the fMRI session. During fMRI, they completed an Anticipation Task (two runs), an Integration Task (four runs, data not included in the current manuscript) and a Localizer Task (four runs). In the training session, stimuli were presented on a computer screen with PsychoPy[62] and in virtual reality with an Oculus Rift and Unity, using a mixture of custom code and OpenMaze[63]. In the fMRI session, PsychoPy was used to present the stimuli, which were projected onto a screen in the scanner bore and viewed via a mirror mounted on the head coil .

Training Phase

In the training phase (one day before the fMRI scan), participants were instructed to learn the order of the four sequences (Map A Green, Map A Blue, Map B Green, Map B Blue; see *Stimuli and Sequence Structure)*. Participants always began by learning the Map A Green Path, because Map A was defined as the first set of environments that participants learned and the Green Path was defined as the first sequence within each map.

Participants were instructed to learn the sequences by generating a story to link the environments in order. They first saw 2D renderings of all the environments in the Map A Green Path order displayed on a computer screen. They were told to generate a detailed story to link the environments in order, and that the final environment should loop back to the first environment in the sequence to create a circle. Participants indicated that they were finished generating a story by pressing a button. Then, they were shown the sequence as pairs of adjacent environments with an empty text entry box displayed underneath (e.g., environments #1 and #2, then environments #2 and #3, etc). Participants were told to write down the story that

they had generated (**Figure 1b**; see **Supplementary Material** for story examples). Participants were given unlimited time to generate and write down their story. Once they had finished, participants verbally repeated the story back to the experimenter.

Following story generation, participants then experienced the Map A Green Path in virtual reality using an Oculus Rift (**Figure 1c**). Participants were initially placed in the first environment in the sequence. After five seconds, a floating green sphere and blue sphere appeared in a random location within reaching distance of the participant. Participants were told that touching the spheres would teleport them to the next environment in the correspondingly colored sequence: they were told to touch the green sphere on the Green Path and the blue sphere on the Blue Path. After being teleported to the next environment in the corresponding sequence, participants were again given five seconds to explore the environment before the spheres would appear. After 20% of trials ("test trials"), instead of teleporting to the next environment in the sequence, participants were teleported to a black environment in which they were shown two images of upcoming environments and were told to indicate which of those two environments was coming up sooner in the sequence they were currently "traversing", relative to the preceding environment. Participants had ten seconds to respond using the Y and B buttons on the left and right Oculus Rift controllers. They were given feedback about whether their answer was correct or incorrect. As participants were exploring the environments in virtual reality, they were also told to rehearse their stories to ensure the sequence was learned. Participants rehearsed the Map A Green Path sequence in virtual reality three times following this procedure.

Participants then repeated the exact same procedure, but learned the Map A Blue Path, which consisted of the same environments as the Map A Green Path in a different order. Participants were told to make their Blue Path story distinct from their Green Path story to avoid confusing the two paths. They then followed the same virtual reality procedure as noted above,

but were instructed to touch the blue spheres instead of the green spheres to teleport between environments.

Following Map A Green and Blue Path learning, participants were exposed to each sequence three more times (including test trials) in virtual reality in an interleaved fashion (i.e., one presentation of Green Path then one presentation of Blue path, repeated three times). Participants then recalled the order of the Map A Green and Blue Paths. The above procedure was then repeated for the Map B Green and Blue Paths. In total, the training phase took between one and a half and two hours to complete. All participants performed at ceiling by the end of the training phase.

Sequence Refresher Task

Participants returned one day later. Before the fMRI scan, they completed a sequence refresher task to ensure they maintained memory for all four sequences learned during the Training Phase. Participants viewed 2D renderings of all the environments from virtual reality, one at a time, in the order of each of the four sequences (Map A Green, Map A Blue, Map B Green, Map B Blue). Participants saw each sequence in order three times. In the first presentation, participants were told to verbally repeat the stories they had generated for each sequence. In the subsequent two presentations, participants were told to verbally recall the environment that came after the currently presented environment in the current sequence.

Anticipation Task

During the fMRI scan, participants first completed the Anticipation Task, for which there were two runs with 32 trials each (**Figure 2a)**. On each trial in the Anticipation task, participants were cued with an environment and a path cue (“Green” or “Blue”) for 3 seconds. This cue indicated the starting point and sequence on that trial. Participants then viewed a blank (gray)

screen for a variable duration (five to nine seconds). Then, participants were presented with two images of upcoming environments and were told to judge which of the two environments was coming up sooner in the cued sequence, relative to the cue image. Participants were given three seconds to make this judgment. This relatively short response deadline was implemented to encourage participants to use the blank screen period to generate predictions along the cued path in preparation for the forced choice decision. The correct answer could be one to four steps away from the cue image. The incorrect answer could be a maximum of five steps away from the cue image. Because the sequences were circular, every environment could be used as a cue with successors up to five steps away. There was a uniformly sampled three to eight second jittered inter-trial interval (ITI), during which participants viewed a fixation cross. At the end of each run, there was a 60 second rest period during which participants viewed a blank screen.

In each run, participants were cued with every environment from Map A and B on the Blue and Green Paths (eight environments per sequence) for a total of 32 trials per run (64 trials total). In the probe phase, the correct answer was equally distributed across steps into the future (one to four). The incorrect answer was randomly sampled to be one to four steps away from the correct answer (two to five steps away from the cue). Within a run, sequences were presented in blocks (i.e., participants completed the Anticipation Task for all the environments in the Map A Green Path in one block), but the order of the cues was randomized within a block. The order of the sequence blocks was also randomized across runs and participants. A single run of the Anticipation Task was approximately 11 minutes, for a total of 22 minutes across both runs.

Integration Task

Following the Anticipation Task, participants completed an Integration Task, in which they were told that one of the environments in Map A was now connected to one of the environments in Map B on either the Green or Blue Path. One of the environments in Map B also connected

back to Map A, creating a single integrated path encompassing all the environments in both maps. The integrated path (Green or Blue) was counterbalanced across participants, with the other path serving as a control non-integrated path. For example, if a participant learned that Map A and Map B were integrated on the Green Path, the Blue Path would be the non-integrated path. The environments that connected Map A to Map B were randomly selected, while the environments that connected Map B back to Map A were always the preceding environments in the sequence, allowing the integrated path to form a circle. Participants then completed a version of the Anticipation Task (see above) in which they anticipated upcoming environments in the non-integrated and integrated paths (four runs, 24 trials per run). The Integration Task is not analyzed in the current manuscript.

Localizer Task

Participants then completed four runs of a Localizer Task used to obtain environment-specific patterns of brain activity across participants (**Figure 3a,** see *Environment Templates).* In the Localizer Task, participants were cued with an environment from Map A or B on the screen for one second. The cue in the Localizer Task did not include a path cue (Green or Blue), allowing us to obtain a context-independent pattern of brain activity for each environment. Specifically, the lack of a context cue should disincentivize participants from consistently activating one sequence (Green or Blue path) over the other while viewing the environments – allowing us to obtain activity patterns for each environment relatively uncontaminated by associated information. Following the cue, participants saw a blank gray screen for five seconds, during which they were told to imagine being inside the environment in virtual reality. Participants then viewed images of the cued environment from different angles, 45 degrees apart, for four seconds. They were then given three seconds to rate how well their imagination matched the

actual images of the environment, on a scale from one to four (one = not well, four = very well). There was a three to eight second jittered ITI, during which participants viewed a fixation cross.

In each run, participants were cued with every environment from Map A and B for a total of 16 trials per run (64 trials total across all four runs). The order of the environments was randomized across runs and participants. A single run of the Localizer Task was approximately five and a half minutes, for a total of 22 minutes across all four runs.

Behavioral Analysis

We conducted analyses on the behavioral data in the R programming language using generalized linear and linear mixed effects models (GLMMs and LMMs, glmer and lmer functions in the *lme4* package[64]). For analyses that modeled multiple observations per participant, such as accuracy or response time on a given trial, models included random intercepts and slopes for all within-participant effects. All response time models examined responses on correct trials only.

To ensure that participants performed effectively during the Anticipation Task, we first tested whether accuracy during the Probe screen (see **Figure 2a**) was better than chance performance (50%) using a one-sample t-test.

We next determined whether accuracy and response time differed across the Maps (A and B) and Paths (Green and Blue). To examine sequence effects, we fit separate models for accuracy (a GLMM) and response time (an LMM) as a function of Map (A = -0.5, B = 0.5), Path (Green = -0.5, Blue = 0.5), and their interaction. We used the following R-based formulas (where "participant" indicates participant number):

*glmer(correct ~ map*path + (1+map*path|participant), family = "binomial", data)*

*lmer(RT ~ map*path + (1+map*path|participant), data, subset = (correct == 1))*

Next, we determined whether accuracy and response time differed across steps into the future. We fit separate models for accuracy (a GLMM) and response time (an LMM) as a function of steps into the future (-0.75 = 1 step, -0.25 = 2 steps, 0.25 = 3 steps, 0.75 = 4 steps). We used the following R-based formulas (where "participant" indicates participant number):

glmer(correct ~ steps + (1+steps|participant), family = "binomial", data)

lmer(RT ~ steps + (1+steps|participant), data, subset = (correct == 1))

MRI Acquisition

Whole-brain data were acquired on a 3 Tesla Siemens Magnetom Prisma scanner equipped with a 64-channel head coil at Columbia University. Whole-brain, high-resolution (1.0 mm iso) T1 structural scans were acquired with a magnetization-prepared rapid acquisition gradient-echo sequence (MPRAGE) at the beginning of the scan session. Functional measurements were collected using a multiband echo-planar imaging (EPI) sequence (repetition time = 1.5s, echo time = 30ms, in-plane acceleration factor = 2, multiband acceleration factor = 3, voxel size = 2mm iso). Sixty-nine oblique axial slices were obtained in an interleaved order. All slices were tilted approximately -20 degrees relative to the AC-PC line. There were ten functional runs in total: two runs of the Anticipation Task, four runs of an Integration Task (not analyzed here), and four runs of the Localizer Task. Field maps were collected after the final functional scan to aid registration (TR = 679 ms, TE = 4.92 ms/7.38 ms, flip angle = 60°, 69 slices, 2 mm isotropic).

Preprocessing

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 1.5.2 (Esteban, Markiewicz, et al. (2018)[65]; Esteban, Blair, et al. (2018)[66]; RRID:SCR_016216), which

is based on *Nipype* 1.3.1 (Gorgolewski et al. (2011)[67]; Gorgolewski et al. (2018)[66]; RRID:SCR_002502).

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010)[68], distributed with ANTs 2.2.0 (Avants et al. 2008[69], RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, RRID:SCR_002823, Zhang, Brady, and Smith 2001[70]). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR_001847, Dale, Fischl, and Sereno 1999[71]), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017[72]). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al. (2009)[73], RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

For each of the 10 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. A deformation field to correct for

susceptibility distortions was estimated based on a field map that was co-registered to the BOLD reference, using a custom workflow of *fMRIPrep* derived from D. Greve's epidewarp.fsl script and further improvements of HCP Pipelines (Glasser et al. 2013)[74]. Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al. 2002)[75]. The BOLD time-series, were resampled to surfaces on the following spaces: *fsaverage6*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in ['MNI152NLin2009cAsym'] space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al. 2014)[76]. The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al. 2007)[77]. Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine

filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964)[78]. Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.5.2 (Abraham et al. 2014[79], RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation.

Copyright Waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the CC0 license.

fMRI Analysis

After preprocessing, all fMRI analyses were performed in Python and R. Pattern similarity analyses were performed using custom code in Python 3. Statistical analysis comparing pattern similarity values across conditions, correlations between fMRI results and behavior, and visualizations were performed using custom code in R.

Localizer Task Analyses

We conducted GLMs predicting whole-brain univariate BOLD activity from task and nuisance regressors from the Localizer Task using custom scripts in Python. For each participant, we first concatenated the fMRI data across runs of the Localizer Task and modeled BOLD activity for each environment (1 to 16) with a boxcar regressor combined across the cue, blank screen, and panorama periods. We also included nuisance regressors in the same model (translation and rotation along the X, Y, and Z axes and their derivatives, motion outliers as determined by fMRIprep, CSF, white matter, framewise displacement, and discrete cosine-basis regressors for periods up to 125 seconds).

We next looked across the whole brain for voxels that showed reliable, environment-specific patterns of activity during the Localizer Task. We used an approach that identifies voxels that respond reliably to different conditions across runs of an experiment[80], here measuring reliability across different participants[81]. For each voxel, we obtained a 16-element

vector of beta weights from the whole-brain GLM, reflecting the beta weight for each of the 16 environments for each participant (e.g., Participant #1 or P1). Next, we obtained the Pearson correlation (r) between each participant's 16-element vector in each voxel and the averaged 16-element vector from the remaining participants (e.g., P2-P32). Finally, we calculated an environment reliability score by averaging the r values across all iterations of the held-out participant (**Figure 3b**). Voxels that had an *r* value of 0.1 or greater ("environment-reliable voxels") were then included in subsequent steps (**Figure 3c**). We selected 0.1 as our cutoff because it resulted in reasonable spatial coverage while maintaining voxel reliability, including in our regions of interest[80].

Conjunction ROI Definition

Three *a priori* regions of interest (ROIs) were defined using environment-reliable voxels (see above) within anatomical or functional areas of interest. The V1-4 ROI was obtained from the probabilistic human visual cortex atlas provided in Wang et al.[82] (threshold: p = 0.50). The hippocampus and insular cortex ROIs were both defined from the Harvard-Oxford probabilistic atlas in FSL (threshold: p = 0.50). We resampled the three ROIs onto the same MNI grid as the functional data (MNI152NLin2009cAsym), and then intersected them with our map of environment-reliable voxels ($r > 0.1$, see description above) to create conjunction ROIs in visual cortex, hippocampus, and insula (**Figure 3d**).

We then obtained the spatial pattern of activity across voxels in each conjunction ROI; these served as environment-specific template activity patterns for each participant. Because the Localizer Task did not include a path cue (Green or Blue), participants should not have been differentially and consistently activating one path as they viewed each environment; thus, the pattern of activity obtained for each environment should be context-independent and should not

prioritize past or upcoming environments in a given context. Importantly, this approach yielded the expected result of producing ROIs with environment-specific patterns of activity: activity patterns for the same environment were more correlated than activity patterns for different environments within each conjunction ROI (**Figure 3d**). These environment-specific patterns (hereafter referred to as "environment templates") are a necessary precursor for investigating prediction along each sequence (see below).

Anticipation Task Analyses

We conducted GLMs predicting whole-brain univariate BOLD activity from behavioral and nuisance regressors from the Anticipation Task using Python. For each participant, we modeled BOLD activity concatenated across both runs of the Anticipation Task with separate regressors for the cue, blank screen, and probe periods for each environment in Map A and B (1 to 16) and for each path (Green Path and Blue Path). This resulted in a total of 32 task regressors for each phase (cue, blank screen, probe) of the Anticipation Task (16 environments across Map A and Map B, with each environment modeled separately for the Green Path and the Blue Path). We also included nuisance regressors in the same model (the same as those used for the Localizer Task Analyses). For all subsequent analyses (except the searchlight analysis), the resulting beta weights were examined within our conjunction ROIs.

Cue Period - Asymmetrical Gaussian Analysis

To assess evidence for multivoxel representations of temporal structure, we obtained the correlation between (1) a given participant's (e.g., P1) cue screen activity pattern for each trial type (a given environment cued on a given path) in the Anticipation Task and (2) the remaining participants' (e.g., P2-P32) averaged patterns of activity for each of the environment templates

from the corresponding map (**Figure 4a**). For example, if Participant 1 was cued with environment one from Map A on the Green Path, we obtained the correlation between: (1) Participant 1's cue screen activity pattern for that environment and path and (2) each environment template (averaged across Participants 2-32) from Map A. This yielded eight separate pattern similarity values (because there are eight environments per map) for each trial type (a given environment cued on a given path). Because each participant had a different ordering of environments, these across-participant templates cannot contain any reliable information about the successors or predecessors of a given environment for a given participant. We then ordered the resulting pattern similarity values according to the cued Map and Path (in this example, the Map A Green Path order), with the cue in the center (**Figure 4a**). Thus, successors following the cue would be to the right of the center and predecessors would be to the left of the center. Because, in an eight-environment map, four steps away is an equal distance from the cue in both the past and the future, we included the pattern similarity value four steps away from the cue in both the forward and the backward direction. We also obtained the correlation between (1) a given participant's (e.g., P1) cue screen activity pattern for each trial type (a given environment cued on a given path) in the Anticipation Task and (2) the remaining participants' (e.g., P2-P32) averaged patterns of activity across all environment templates in the different map (in this example, Map B). This single value served as the different-map baseline.

Next, we fit an asymmetrical Gaussian curve to the resulting ordered pattern similarity values. We chose to use a Gaussian curve because we hypothesized that brain regions would represent upcoming (or past) environments in a graded manner, with stronger representations for nearby environments[23]. The asymmetrical Gaussian has four parameters: amplitude, asymptote, and forward and backward widths (σ) (**Figure 4b**). The amplitude controls the height of the peak of the Gaussian curve, and indicates the extent to which a brain region is representing the cue

environment presented on the screen. The asymptote controls the vertical shift of the Gaussian curve; a negative asymptote reflects suppression of some environments (i.e. the cue pattern is anticorrelated with some environment templates). The widths (σ) control the slope of the fall off from the amplitude to the asymptote. Wider Gaussians indicate activation of environment patterns further away from the cue. Because we fit an asymmetrical Gaussian, we obtained different widths in the forward and backward direction; this allows brain regions to potentially represent more environments in one direction (e.g., upcoming environments) than another (e.g., past environments). This in turn enables us to detect if some brain areas anticipate the future but do not represent the past. We constrained the widths to be a maximum of 10 and applied L2 regularization to the amplitude and intercept (with strength = 0.01) to ensure the model did not return uninterpretable parameter values.

To test whether the parameters of the Gaussian curve were consistent across participants, we fit the asymmetrical Gaussian curve on all but one participant's data (e.g. P2-32) and then measured the sum of squared errors (observed vs. predicted pattern similarity values) when using this curve to predict the held-out participant's data (e.g. P1). We repeated this procedure for each choice of held-out participant to obtain an average error value. We compared this error to a null distribution, fitting the asymmetrical Gaussian curve to data in which the order of the environments was shuffled. We defined a p value as the fraction of 10,000 shuffles which produced lower error than our original unshuffled fit.

We also performed statistical tests on the parameters of the correctly ordered Gaussians. We tested whether the amplitude was significantly above the different-map baseline, and whether the asymptote was suppressed below the different-map baseline, using one-sample t-tests across participants. We also tested whether the widths (σ) differed by brain region (visual

cortex vs hippocampus), direction (forward vs backward), and their interaction using the following R-based formula (where “participant” indicates participant number):

*lmer(*width *~ region*direction + (1|participant), data)*

Cue Period - Searchlight

We conducted a whole-brain searchlight analysis with custom Python code to test whether brain regions beyond our ROIs represented temporal structure in the hypothesized asymmetrical Guassian format. We looked for significant Gaussian representations in cubes with a side length of 7 voxels, moved throughout the whole brain volume with a step size of 2 voxels. We included only environment-reliable voxels within each cube, and only proceeded with the analysis of a cube if it contained at least 64 environment-reliable voxels. The parameters of fitted Gaussians within each searchlight, along with the goodness of fit, were assigned to each voxel in the searchlight. For voxels that were included in more than one searchlight, the final Gaussian parameters and goodness-of-fit were obtained by averaging the results across all the searchlights in which the voxel was included.

To determine which voxels exhibited significant Gaussian representations across participants, we first obtained a measure of goodness of fit by dividing the squared errors of the correctly ordered Gaussian by the average of the squared errors of the permuted Gaussians and then subtracting the resulting value from 1 for each voxel included in the searchlight for each participant. Numbers above 0 indicate better fits to the correctly ordered vs permuted data. We then statistically tested whether the goodness of fit values were greater than 0 in each voxel, using FSL’s randomise function with threshold free cluster enhancement, which generates null distributions using 10,000 permutations and performs a one-sample t-test while enhancing

clusters of significant voxels[83]. We then corrected for multiple comparisons using the family-wise error rate correction ($p < 0.05$).

To determine whether Gaussian widths (σ) were organized hierarchically within brain regions, we first averaged the forward and backward widths for each voxel for each participant. We opted to compute the average of the forward and backward widths because we did not find evidence for directional asymmetry in any brain region, in both our ROI and searchlight analyses. Next, we determined whether the averaged widths became increasingly wider in more anterior, compared to posterior, voxels in the parahippocampal place area (PPA) and the retrosplenial cortex (RSC). We created PPA and RSC ROIs using pre-defined anatomical ROIs[84], which we then resampled onto the same MNI grid as the functional data (MNI152NLin2009cAsym) and intersected with our map of environment-reliable voxels ($r > 0.1$). We chose PPA and RSC because (1) the searchlight revealed significant Gaussian representations in the majority of voxels in these regions and (2) they have previously been implicated in both scene perception and memory[36,38]. In each region, for each participant, we obtained the Spearman rank-order correlation between the averaged forward and backward widths (σ) and the y coordinate (indicating a voxel's position on the posterior-anterior axis) across voxels. Finally, we determined whether the correlation was significant at the group level by comparing the participant-specific r values to 0 using a one-sample t-test. A significantly positive r value would indicate that Gaussian curves become increasingly wider in successively anterior aspects of regions.

Cue Period - Relationship to Behavior

We determined whether an individual's asymptote from their Gaussian model, indiciating suppression of environments not captured by the Gaussian's width, was related to response time costs for further environments. Response time costs were quantified with participant-specific

regressions that predicted response time as a function of steps into the future. We then performed an individual differences analysis by obtaining the Spearman rank-order correlation between participants' response time costs and their asymptotes in (1) hippocampus and (2) visual cortex.

Cue Period - Context Dependent Analysis

To test whether multivoxel patterns of activity were context-dependent, we obtained the correlation between a given participant's (e.g., P1) cue screen activity pattern for each trial type (a given environment cued on a given path) in the Anticipation Task and the remaining participants' (e.g., P2-P32) averaged patterns of activity for each of the environment templates (1) one and two steps away from the cue, (2) on the cued path and the uncued path, and (3) in the forward and backward direction (**Figure 7a**). This yielded eight pattern similarity values for each combination of step (one vs two), context (cued vs uncued), and direction (forward vs backward). For each step in each direction, we then obtained a measure of context-dependent representations by subtracting the pattern similarity values for environments on the uncued path from the pattern similarity values for environments on the cued path. For example, context-dependent representations for one step in the forward direction would be calculated as pattern similarity between the cued environment and the environment one step in the forward direction on the *cued* path minus the pattern similarity between the cued environment and the environment one step in the forward direction on the *uncued* path. Positive values indicate prioritized representations for environments on the cued path, relative to the uncued path, whereas negative values indicate suppressed representations for environments on the cued path relative to the uncued path. We then statistically tested whether, at the group-level, context dependence varied as a function of direction (forward vs backward), step (1 step vs 2 step), and their

interaction. We conducted these tests separately for each region (visual cortex, hippocampus, and insula) using the following R-based formula, where cueContextDiff indicates the cued minus uncued pattern similarity values:

*lmer(cueContextDiff ~ direction*step+(1+direction+step|participant), data, subset = region)*

We then tested whether there was context-dependent prioritization or suppression by comparing the cueContextDiff values to 0 with a one-sample t-test.

Blank Period - Context Dependent Analysis

We computed the same context-dependent analyses as described above, but using the blank screen activity pattern, rather than the cue screen, for each trial type (a given environment cued on a given path) of the Anticipation Task (**Figure 7a**). We hypothesized that we would be more likely to find one step, context-dependent representations in visual cortex during the blank screen period compared to the cue screen because (1) this region is strongly modulated by visual input, potentially reducing our ability to detect anticipatory representations while images are on the screen and (2) past work has shown short timescale (e.g., one step) predictions in this region[15,16]. Based on these hypotheses, we statistically tested whether, at the group-level, visual cortex exhibited prioritized or suppressed context-dependent representations one step into the future in the forward direction using a one-sample t-test.

Declaration of Interests

The authors declare no competing interests.

References

1. Tolman, E.C. (1948). Cognitive maps in rats and men. Psychol. Rev. *55*, 189–208. 10.1037/h0061626.
2. Stachenfeld, K.L., Botvinick, M.M., and Gershman, S.J. (2017). The hippocampus as a predictive map. Nat. Neurosci. *20*, 1643–1653. 10.1038/nn.4650.
3. Behrens, T.E.J., Muller, T.H., Whittington, J.C.R., Mark, S., Baram, A.B., Stachenfeld, K.L., and Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. Neuron *100*, 490–509. 10.1016/j.neuron.2018.10.002.
4. Momennejad, I. (2020). Learning Structures: Predictive Representations, Replay, and Generalization. Curr. Opin. Behav. Sci. *32*, 155–166. 10.1016/j.cobeha.2020.02.017.
5. Buckner, R.L. (2010). The Role of the Hippocampus in Prediction and Imagination. Annu. Rev. Psychol. *61*, 27–48. 10.1146/annurev.psych.60.110707.163508.
6. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. *36*, 181–204. 10.1017/S0140525X12000477.
7. Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. B Biol. Sci. *360*, 815–836. 10.1098/rstb.2005.1622.
8. Summerfield, C., and Egner, T. (2009). Expectation (and attention) in visual cognition. Trends Cogn. Sci. *13*, 403–409. 10.1016/j.tics.2009.06.003.
9. Lisman, J., and Redish, A. d. (2009). Prediction, sequences and the hippocampus. Philos. Trans. R. Soc. B Biol. Sci. *364*, 1193–1201. 10.1098/rstb.2008.0316.
10. Ouden, H.E.M. den, Daunizeau, J., Roiser, J., Friston, K.J., and Stephan, K.E. (2010). Striatal Prediction Error Modulates Cortical Coupling. J. Neurosci. *30*, 3210–3219. 10.1523/JNEUROSCI.4458-09.2010.
11. de Lange, F.P., Heilbron, M., and Kok, P. (2018). How Do Expectations Shape Perception? Trends Cogn. Sci. *22*, 764–779. 10.1016/j.tics.2018.06.002.
12. Lee, C.S., Aly, M., and Baldassano, C. (2021). Anticipation of temporally structured events in the brain. eLife *10*, e64972. 10.7554/eLife.64972.
13. Kok, P., Jehee, J.F.M., and de Lange, F.P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. Neuron *75*, 265–270. 10.1016/j.neuron.2012.04.034.
14. Kok, P., Failing, M.F., and de Lange, F.P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. J. Cogn. Neurosci. *26*, 1546–1554. 10.1162/jocn_a_00562.
15. Gavornik, J.P., and Bear, M.F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. Nat. Neurosci. *17*, 732–737. 10.1038/nn.3683.
16. Hindy, N.C., Ng, F.Y., and Turk-Browne, N.B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nat. Neurosci. *19*, 665–667. 10.1038/nn.4284.
17. Schapiro, A.C., Kustner, L.V., and Turk-Browne, N.B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. Curr. Biol. *22*, 1622–1627. 10.1016/j.cub.2012.06.056.
18. Davachi, L., and DuBrow, S. (2015). How the hippocampus preserves order: the role of prediction and context. Trends Cogn. Sci. *19*, 92–99. 10.1016/j.tics.2014.12.004.
19. Brown, T.I., Carr, V.A., LaRocque, K.F., Favila, S.E., Gordon, A.M., Bowles, B., Bailenson, J.N., and Wagner, A.D. (2016). Prospective representation of navigational goals in the human hippocampus. Science *352*, 1323–1326. 10.1126/science.aaf0784.
20. Kok, P., and Turk-Browne, N.B. (2018). Associative Prediction of Visual Shape in the Hippocampus. J. Neurosci. *38*, 6888–6899. 10.1523/JNEUROSCI.0163-18.2018.

21. Aitken, F., and Kok, P. (2022). Hippocampal representations switch from errors to predictions during acquisition of predictive associations. Nat. Commun. *13*, 3294. 10.1038/s41467-022-31040-w.
22. Brunec, I.K., and Momennejad, I. (2022). Predictive Representations in Hippocampal and Prefrontal Hierarchies. J. Neurosci. *42*, 299–312. 10.1523/JNEUROSCI.1327-21.2021.
23. Momennejad, I., and Howard, M.W. (2018). Predicting the Future with Multi-scale Successor Representations. 449470. 10.1101/449470.
24. Hasson, U., Chen, J., and Honey, C.J. (2015). Hierarchical process memory: memory as an integral component of information processing. Trends Cogn. Sci. *19*, 304–313. 10.1016/j.tics.2015.04.006.
25. Hasson, U., Yang, E., Vallines, I., Heeger, D.J., and Rubin, N. (2008). A Hierarchy of Temporal Receptive Windows in Human Cortex. J. Neurosci. *28*, 2539–2550. 10.1523/JNEUROSCI.5487-07.2008.
26. Lerner, Y., Honey, C.J., Silbert, L.J., and Hasson, U. (2011). Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. J. Neurosci. *31*, 2906–2915. 10.1523/JNEUROSCI.3684-10.2011.
27. Baldassano, C., Chen, J., Zadbood, A., Pillow, J.W., Hasson, U., and Norman, K.A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. Neuron *95*, 709-721.e5. 10.1016/j.neuron.2017.06.041.
28. Aly, M., Chen, J., Turk-Browne, N.B., and Hasson, U. (2018). Learning Naturalistic Temporal Structure in the Posterior Medial Network. J. Cogn. Neurosci. *30*, 1345–1365. 10.1162/jocn_a_01308.
29. Foster, D.J., and Wilson, M.A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature *440*, 680–683. 10.1038/nature04587.
30. Diba, K., and Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. Nat. Neurosci. *10*, 1241–1242. 10.1038/nn1961.
31. Jadhav, S.P., Kemere, C., German, P.W., and Frank, L.M. (2012). Awake Hippocampal Sharp-Wave Ripples Support Spatial Memory. Science *336*, 1454–1458. 10.1126/science.1217230.
32. Ólafsdóttir, H.F., Bush, D., and Barry, C. (2018). The Role of Hippocampal Replay in Memory and Planning. Curr. Biol. *28*, R37–R50. 10.1016/j.cub.2017.10.073.
33. Wimmer, G.E., Liu, Y., Vehar, N., Behrens, T.E.J., and Dolan, R.J. (2020). Episodic memory retrieval success is associated with rapid replay of episode content. Nat. Neurosci. *23*, 1025–1033. 10.1038/s41593-020-0649-z.
34. Crivelli-Decker, J., Clarke, A., Park, S.A., Huffman, D.J., Boorman, E.D., and Ranganath, C. (2023). Goal-oriented representations in the human hippocampus during planning and navigation. Nat. Commun. *14*, 2946. 10.1038/s41467-023-35967-6.
35. Tarder-Stoll, H., Baldassano, C., and Aly, M. (2023). Consolidation enhances multistep anticipatory judgements but diminishes access to perceptual features. 10.31234/osf.io/x2f7s.
36. Epstein, R.A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn. Sci. *12*, 388–396. 10.1016/j.tics.2008.07.004.
37. Steel, A., Billings, M.M., Silson, E.H., and Robertson, C.E. (2021). A network linking scene perception and spatial memory systems in posterior cerebral cortex. Nat. Commun. *12*, 2632. 10.1038/s41467-021-22848-z.
38. Baldassano, C., Beck, D.M., and Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. NeuroImage *75*, 228–237. 10.1016/j.neuroimage.2013.02.073.
39. Dayan, P. (1993). Improving Generalization for Temporal Difference Learning: The Successor Representation. Neural Comput. *5*, 613–624. 10.1162/neco.1993.5.4.613.

40. Momennejad, I., Russek, E.M., Cheong, J.H., Botvinick, M.M., Daw, N.D., and Gershman, S.J. (2017). The successor representation in human reinforcement learning. Nat. Hum. Behav. *1*, 680–692. 10.1038/s41562-017-0180-8.
41. Ekman, M., Kusch, S., and de Lange, F.P. (2023). Successor-like representation guides the prediction of future events in human visual cortex and hippocampus. eLife *12*, e78904. 10.7554/eLife.78904.
42. Deuker, L., Bellmund, J.L., Navarro Schröder, T., and Doeller, C.F. (2016). An event map of memory space in the hippocampus. eLife *5*, e16534. 10.7554/eLife.16534.
43. Fernandez, C., Jiang, J., Wang, S.-F., Choi, H.L., and Wagner, A.D. (2023). Representational integration and differentiation in the human hippocampus following goal-directed navigation. eLife *12*, e80281. 10.7554/eLife.80281.
44. Elliott Wimmer, G., and Büchel, C. (2019). Learning of distant state predictions by the orbitofrontal cortex in humans. Nat. Commun. *10*, 2554. 10.1038/s41467-019-10597-z.
45. Turk-Browne, N.B., Simon, M.G., and Sederberg, P.B. (2012). Scene Representations in Parahippocampal Cortex Depend on Temporal Context. J. Neurosci. *32*, 7202–7207. 10.1523/JNEUROSCI.0942-12.2012.
46. Aminoff, E., Gronau, N., and Bar, M. (2007). The Parahippocampal Cortex Mediates Spatial and Nonspatial Associations. Cereb. Cortex *17*, 1493–1503. 10.1093/cercor/bhl078.
47. Bar, M., Aminoff, E., and Schacter, D.L. (2008). Scenes Unseen: The Parahippocampal Cortex Intrinsically Subserves Contextual Associations, Not Scenes or Places Per Se. J. Neurosci. *28*, 8539–8544. 10.1523/JNEUROSCI.0987-08.2008.
48. Aminoff, E.M., Kveraga, K., and Bar, M. (2013). The role of the parahippocampal cortex in cognition. Trends Cogn. Sci. *17*, 379–390. 10.1016/j.tics.2013.06.009.
49. Marchette, S.A., Vass, L.K., Ryan, J., and Epstein, R.A. (2015). Outside Looking In: Landmark Generalization in the Human Navigational System. J. Neurosci. *35*, 14896–14908. 10.1523/JNEUROSCI.2270-15.2015.
50. Clarke, A., Crivelli-Decker, J., and Ranganath, C. (2022). Contextual Expectations Shape Cortical Reinstatement of Sensory Representations. J. Neurosci. *42*, 5956–5965. 10.1523/JNEUROSCI.2045-21.2022.
51. Davachi, L. (2006). Item, context and relational episodic encoding in humans. Curr. Opin. Neurobiol. *16*, 693–700. 10.1016/j.conb.2006.10.012.
52. Eichenbaum, H., Yonelinas, A.P., and Ranganath, C. (2007). The Medial Temporal Lobe and Recognition Memory. Annu. Rev. Neurosci. *30*, 123–152. 10.1146/annurev.neuro.30.051606.094328.
53. Schlichting, M.L., Mumford, J.A., and Preston, A.R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat. Commun. *6*, 8151. 10.1038/ncomms9151.
54. Duncan, K.D., and Schlichting, M.L. (2018). Hippocampal representations as a function of time, subregion, and brain state. Neurobiol. Learn. Mem. *153*, 40–56. 10.1016/j.nlm.2018.03.006.
55. Dimsdale-Zucker, H.R., Ritchey, M., Ekstrom, A.D., Yonelinas, A.P., and Ranganath, C. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. Nat. Commun. *9*, 294. 10.1038/s41467-017-02752-1.
56. Brunec, I.K., Robin, J., Olsen, R.K., Moscovitch, M., and Barense, M.D. (2020). Integration and differentiation of hippocampal memory traces. Neurosci. Biobehav. Rev. *118*, 196–208. 10.1016/j.neubiorev.2020.07.024.
57. Molitor, R.J., Sherrill, K.R., Morton, N.W., Miller, A.A., and Preston, A.R. (2021). Memory Reactivation during Learning Simultaneously Promotes Dentate Gyrus/CA2,3 Pattern Differentiation and CA1 Memory Integration. J. Neurosci. *41*, 726–738.

10.1523/JNEUROSCI.0394-20.2020.
58. Bein, O., and Davachi, L. (2022). Event integration and temporal pattern separation: how hierarchical knowledge emerges in hippocampal subfields through learning. 2022.07.18.500527. 10.1101/2022.07.18.500527.
59. Manning, J.R., Polyn, S.M., Baltuch, G.H., Litt, B., and Kahana, M.J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. Proc. Natl. Acad. Sci. *108*, 12893–12897. 10.1073/pnas.1015174108.
60. DuBrow, S., and Davachi, L. (2014). Temporal Memory Is Shaped by Encoding Stability and Intervening Item Reactivation. J. Neurosci. *34*, 13998–14005. 10.1523/JNEUROSCI.2535-14.2014.
61. Clewett, D., and Davachi, L. (2017). The ebb and flow of experience determines the temporal structure of memory. Curr. Opin. Behav. Sci. *17*, 186–193. 10.1016/j.cobeha.2017.08.013.
62. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods *51*, 195–203. 10.3758/s13428-018-01193-y.
63. Alsbury-Nealy, K., Wang, H., Howarth, C., Gordienko, A., Schlichting, M.L., and Duncan, K.D. (2022). OpenMaze: An open-source toolbox for creating virtual navigation experiments. Behav. Res. Methods *54*, 1374–1387. 10.3758/s13428-021-01664-9.
64. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. 10.48550/arXiv.1406.5823.
65. Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods *16*, 111–116. 10.1038/s41592-018-0235-4.
66. Esteban, O., Markiewicz, C.J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., Berleant, S., Ellis, D.G., Pinsard, B., Madison, C., et al. (2022). nipy/nipype: 1.8.3. 10.5281/zenodo.6834519.
67. Gorgolewski, K., Burns, C., Madison, C., Clark, D., Halchenko, Y., Waskom, M., and Ghosh, S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. Front. Neuroinformatics *5*.
68. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., and Gee, J.C. (2010). N4ITK: Improved N3 Bias Correction. IEEE Trans. Med. Imaging *29*, 1310–1320. 10.1109/TMI.2010.2046908.
69. Avants, B.B., Epstein, C.L., Grossman, M., and Gee, J.C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. *12*, 26–41. 10.1016/j.media.2007.06.004.
70. Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging *20*, 45–57. 10.1109/42.906424.
71. Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage *9*, 179–194. 10.1006/nimg.1998.0395.
72. Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E.C., et al. (2017). Mindboggling morphometry of human brains. PLOS Comput. Biol. *13*, e1005350. 10.1371/journal.pcbi.1005350.
73. Fonov, V., Evans, A., McKinstry, R., Almli, C., and Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage *47*, S102. 10.1016/S1053-8119(09)70884-5.
74. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage *80*, 105–124. 10.1016/j.neuroimage.2013.04.127.

75. Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage *17*, 825–841. 10.1006/nimg.2002.1132.
76. Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage *84*, 320–341. 10.1016/j.neuroimage.2013.08.048.
77. Behzadi, Y., Restom, K., Liau, J., and Liu, T.T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. NeuroImage *37*, 90–101. 10.1016/j.neuroimage.2007.04.042.
78. Lanczos, C. (1964). Evaluation of Noisy Data. J. Soc. Ind. Appl. Math. Ser. B Numer. Anal. *1*, 76–85. 10.1137/0701007.
79. Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Front. Neuroinformatics *8*.
80. Tarhan, L., and Konkle, T. (2020). Reliability-based voxel selection. NeuroImage *207*, 116350. 10.1016/j.neuroimage.2019.116350.
81. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. Science *303*, 1634–1640. 10.1126/science.1089506.
82. Wang, L., Mruczek, R.E.B., Arcaro, M.J., and Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. Cereb. Cortex *25*, 3911–3931. 10.1093/cercor/bhu277.
83. Smith, S.M., and Nichols, T.E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage *44*, 83–98. 10.1016/j.neuroimage.2008.03.061.
84. Julian, J.B., Fedorenko, E., Webster, J., and Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. NeuroImage *60*, 2357–2364. 10.1016/j.neuroimage.2012.02.055.