Title
Characterizing the Nature of Goal Representations in Memory

Author
Crivelli-Decker, Jordan

Publication Date
2022

Peer reviewed|Thesis/dissertation

Characterizing the Nature of Goal Representations in Memory

By

JORDAN CRIVELLI-DECKER

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Charan Ranganath, Chair

Randall O'Reilly

Erie D. Boorman

Committee in Charge

2022

Table of Contents

Abstract

Theories of hippocampal function have consistently focused on this remarkable regions' important role in spatial navigation and memory. Recent work in cognitive and systems neuroscience have suggested that the hippocampus might support planning, imagination, and navigation by forming "cognitive maps" that capture the abstract structure of physical spaces, tasks, and situations. However, a critical aspect of planning and navigation is that both involve simulating, and acting upon, a sequence of actions to reach a goal. Despite several decades of work investigating hippocampal function, the precise mechanism and neural basis of planning during navigation is still unclear. This dissertation investigated how goals impact activity patterns in the hippocampus during planning and navigation. In Part 1, we examined hippocampal activity patterns in humans, using a goal-directed navigation task, to examine how goal information is incorporated in the construction and execution of navigational plans. Interestingly, we found that hippocampal activity patterns were more similar when participants planned routes that led to the same goal. During navigation, we found that rather than simply representing the current location in space, or the immediate future (as predicted by many theories of hippocampal function), the hippocampus reactivated a key decision point along a route. Building on these findings, in Part 2, a biologically inspired neural network of the hippocampus was used to investigate two plausible theories of the mechanisms underlying planning in humans; chaining and chunking. To simulate chaining during planning, the network was trained to retrieve pairs of overlapping associations in sequence. Using this framework, the model performed quite poorly at moments of high overlap between other sequences. This was in contrast to a

model that utilized chunking, which was provided with only the most relevant information for retrieving a navigational memory. Taken together, this dissertation extends prior work in planning and navigation by highlighting the importance of external inputs into the hippocampal circuit to provide both structure and goal-relevant information, which are essential components of human memory.

Acknowledgments

To every mentor, professor, family member, friend, and colleague who has encouraged me to pursue higher education and kindled my passion for intellectual discovery; thank you.

I would like to thank my graduate advisor, Dr. Charan Ranganath, for his unwavering support and mentorship throughout my academic career. Thank you for hiring me as a lab manager and giving me the opportunity to discover one of the greatest passions of my life. Your mentorship has helped me become a better scientist, guided me through adulthood, and taught me how to make tough decisions. Through the majority of my 20's I spent my time investing in myself in your lab and I will be forever grateful for this opportunity.

I would also like to thank my dissertation committee members. Erie Boorman and Randall O'Reilly. Thank you, Erie, for the many hours discussing possible experiments and allowing me to attend your lab meetings. Your guidance helped me discover new avenues for academic research. Randy, I appreciate you taking the time to have long meetings with me going over code, finding bugs, and discussing science. I would also like to thank additional faculty members whose mentorship had a formative impact on me throughout my PhD – Andrew Fox, Andrew Yonelinas, Philippe Rast, Mark Goldman, and Jeremey Manning.

Thank you to all of the amazing and brilliant people who have come and gone in the Ranganath lab who made my PhD so much more enjoyable: Brendan Cohn-Sheehy, Walter Reilly, Yicong (Alan) Zheng, Ashley Williams, Shao-Fang (Pam) Wang, Nichole Bouffard, Frank (Liang-Tien) Hsieh, Laura Libby, Marika Inhoff, Halle Dimsdale-Zucker, Tanya Jonker, Zach Reagh, Maureen Ritchey, Matthias Gruber, Eda Mizrak, Alex Clarke, James Antony, Alex Clarke, Alex Park, and many more. I have learned so much from each and every one of you.

To my graduate cohort friends, Brendan Cohn-Sheehy, Kyle Puhger, Mackenzie Englund, and Walter Reilly – I am extremely grateful that I had the chance to cross paths with each of you. You have all become some of my closest friends and I cannot wait to see what all

of you will accomplish in life. I will always look back on my PhD and remember the amazing memories we made at weddings, jam sessions, and backpacking adventures.

To my friends and family, thank you for your love and support throughout this journey. Thank you for listening to me vent countless times and helping me navigate times of transition. To the BH, being able to escape academia and just hang out with you all is one of the only ways I was truly able to get my mind off of work. The whacky and crazy times we have had will always be close to my heart and give me strength during hard times.

To my partner and love of my life, Aileen. Thank you for your never-ending love and patience. I would not be the person I am today without you. Through this crazy journey you have always been there as a steadfast and calming presence. Thank you for always being there to listen to my crazy ideas and supporting me in every way possible. I love you.

Part 1:

Goal-oriented representations in the human hippocampus during planning and navigation

Goal-oriented representations in the human hippocampus during planning and navigation

Jordan Crivelli-Decker[1,2*], Alex Clarke[4], Seongmin A. Park[1,3], Derek J. Huffman[1,5], Erie Boorman [1,3], Charan Ranganath[1,2]

1. Center for Neuroscience, University of California, Davis, USA
2. Department of Psychology, University of California, Davis, USA
3. Center for Mind and Brain, University of California, Davis, USA
4. Department of Psychology, University of Cambridge, UK
5. Department of Psychology, Colby College, USA

* Corresponding Author: Jordan Crivelli-Decker, Email: jecrivellidecker@ucdavis.edu

Abstract

Recent work in cognitive and systems neuroscience has suggested that the hippocampus might support planning, imagination, and navigation by forming “cognitive maps” that capture the abstract structure of physical spaces, tasks, and situations. Navigation involves disambiguating similar contexts, and the planning and execution of a sequence of decisions to reach a goal. We examined hippocampal activity patterns in humans during a goal-directed navigation task to examine how contextual and goal information are incorporated in the construction and execution of navigational plans. During planning, hippocampal pattern similarity was enhanced across routes that shared a context and a goal. This effect could not be explained by stimulus or spatial information alone. During navigation, we observed prospective retrieval of goal-specific hippocampal representations of the key decision point. These results suggest that, rather than simply representing overlapping associations or state transitions, hippocampal activity patterns are shaped by context and goals.

Introduction

Every day, people need to plan and execute actions in order to get what they want. Spatial navigation, for instance, requires one to pull up a mental representation of the relationships between different places—i.e., a “cognitive map” (Tolman 1948)—and generate a plan for how to reach a goal. Tolman (1948) proposed that cognitive maps enable behavioral flexibility, so that the same underlying representation can be used to

reach different goals. For example, if we wanted to navigate to the Tiger exhibit at the San Diego Zoo we might use the same map-like representation to find the Zebra exhibit.

Several lines of evidence suggest that the hippocampus plays a key role in navigation, though its role in navigation is fundamentally unclear. For example, based on findings showing that hippocampal “place cells” encode specific locations within a spatial context, many have argued that the hippocampus forms a cognitive map of physical space (O’Keefe and Dostrovsky, 1971; O’Keefe and Nadel, 1978). It is now clear that the hippocampus also tracks distances in abstract state spaces (Tavares et al., 2015; Park et al., 2019; Aronov et al., 2017), potentially supporting the broader idea that the hippocampus encodes a “memory space” (Eichenbaum and Cohen, 2014) that maps the systematic relationships between any behaviorally relevant variables (Behrens et al 2018, Stachenfeld et al., 2017, Kaplan, Schuck, and Doeller 2017; but see O’Reilly et al. 2022 and Summerfield et al., 2020 for alternative views).

Building on this idea, some have proposed that the hippocampus encodes a “predictive map” that specifies not only one’s *current* location, but also states or locations that could be encountered in the future (e.g. Mehta et al., 2001, Stachenfeld et al., 2017). For example, the “successor representation,” a popular computational implementation of the predictive map model (e.g. Stachenfeld et al., 2017; Gershman, 2018), has been used to argue that the hippocampus represents each state in terms of its possible transitions to future states. This model demonstrates that via an incremental learning process about state-to-state transitions, analogous to model-free learning about rewards, enables organisms to rapidly learn how a sequence of actions can lead to a desired outcome.

Although numerous studies have investigated representations of abstract state spaces in the human hippocampus, two fundamental questions remain unanswered. One key issue concerns the role of context. Single-unit recording studies have reported that the spatial selectivity of place cells is context-specific—that is, the spatial selectivity of a given cell in one environment varies when an animal is moved to a different, but topographically similar environment (O'Keefe and Dostrovsky, 1971, Muller & Kubie, 1987; Skaggs and McNaughton, 1998; Leutgeb et al., 2004, Alme et al., 2014, McKenzie et al., 2014; see Kubie et al., 2020 for review). Just as one might pull up different cognitive maps for different physical contexts, it is reasonable to think that we might utilize context-specific maps of abstract state spaces. Computational models have been proposed to explain how the hippocampus might recognize contexts (Honi et al., 2020, Whittington et al., 2020, George et al., 2021), but there is little empirical evidence showing whether or how context in abstract spaces is encoded by the hippocampus.

A second key issue that has not been addressed is the importance of goals in hippocampal representations of abstract task states. Theories of state space representation by the hippocampus rely heavily on results from studies that examined activity in hippocampal place cells during random movements through an environment (e.g. Alme et al., 2014). Accordingly, studies of abstract spaces in humans typically investigate incidental learning of stimulus dimensions or arbitrary state dynamics (Garvert et al., 2017, Schapiro et al., 2016, Schuck and Niv, 2016). These kinds of passive, incidental learning tasks differ from those used by Tolman (1948) to demonstrate that animals actively use a spatial representation to guide navigation to particular goal locations in an environment. If the human hippocampus forms an

abstract cognitive or predictive map, one would expect to see such a representation during planning and navigation towards different goals in the same context.

Based on what is known from studies of spatial navigation, there is reason to think that hippocampal representations in the context of goal-directed navigation might fundamentally differ from what is seen during random or incidental behavior. For example, hippocampal place cells have differential firing fields during planning depending on the future goal of the animal (Ainge et al., 2007; Wood et al., 2000; Ferbinteanu and Shapiro, 2003, Ito et al., 2015), and goal locations tend to be overrepresented (Dupret et al., 2010, Gauthier et al., 2018). Consistent with these findings, fMRI studies of spatial navigation have found that hippocampal activity is modulated by a participant's distance from a goal location (Patai et al., 2019, Howard et al., 2014), and that hippocampal activity patterns during route planning carry information about prospective goal locations in a virtual space (Brown et al., 2016). These findings suggest that hippocampal representations during planning or navigation in abstract state spaces might be powerfully shaped by goals. If this is indeed the case, it would potentially challenge models proposing that the hippocampus encodes a relatively static map of current (O'Keefe and Dostrovsky, 1971) or possible future states (Stachenfeld et al., 2017).

In the present study, we used functional magnetic resonance imaging (fMRI) to investigate how contexts and goals shape hippocampal representations during planning and navigation (Fig. 1). We devised a task in which participants were required to generate a plan and navigate through two abstract state-space contexts in order to reach a goal state. Critically, the contexts included the same stimuli, with different action

relationships in each context. This allowed us to examine the impact of context and goals during planning and navigation across perceptually similar sequences. We compared activity patterns elicited during planning of sequences that shared a goal to those that had different goals to disentangle the unique contribution of goal information on hippocampal activity patterns. Finally, we analyzed the time course of hippocampal patterns while participants actively navigated during the task to examine if current and future states were reactivated in a way that is consistent with computational models of hippocampal function.

Results

Navigating an abstract spatiotemporal map

Prior to scanning, participants were trained to criterion (85% accuracy) to navigate to four goal animals in two distinct contexts that consisted of animals that were systematically linked in a deterministic sequence structure (see Methods). Each zoo context consisted of the same nine animals arranged in a “plus maze” topology, but the relationships between animals across the two zoos were mirror-reversed and then rotated counterclockwise by 90 degrees (Fig. 1a). At each animal, participants were able to make one of four button presses that allowed them to transition between animals. In the scanner, participants were asked to use their knowledge of the zoo contexts to actively navigate from a "start animal” to a “goal animal” (Fig. 1b), where start and goal animals were always at the ends of the maze arms. Each trial consisted of a planning phase and a navigation phase. During the planning phase, a cue indicated the start and goal animals. Next, during the navigation phase, participants saw the start

animal alone before moving through a sequence of animals to reach the goal animal. For each animal, participants had to decide which direction in the plus maze to move to ultimately reach the goal animal. On any given trial, participants were only allowed four moves to navigate to the goal animal and the interstimulus interval was fixed to ensure that an equal amount of time was spent at each state. In each zoo context, participants planned and navigated 12 distinct sequences (each repeated 4 times across 6 runs of scanning). In addition, one trial from each sequence was randomly chosen to end early at the rabbit (Catch Trials). This resulted in 72 sequences that could be analyzed (see Methods).

Participants were highly accurate at navigating to the goal animal in each context (Context 1: Mean = 93.7%, SD = 12.9%, Context 2: Mean = 94.7%, SD = 12.2%), with no significant differences in accuracy between contexts ($t_{22} = 1.16$, $p = 0.26$). This suggests that participants had successfully formed distinct representations of each zoo

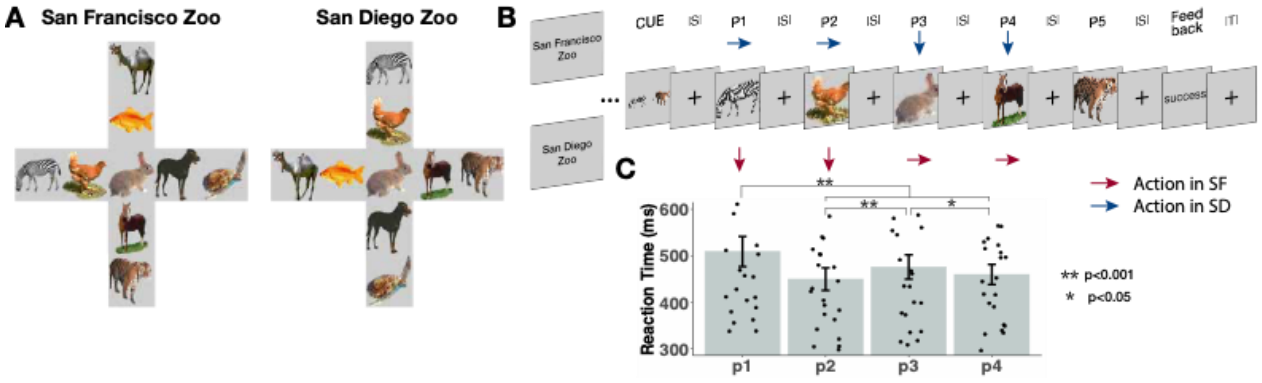

Fig. 1. Task Design and Behavioral Results. A) Overhead view of virtual environments. Each context had the same visual information but the specific spatial orientation was mirror reversed and then rotated counter clockwise 90 degrees. This manipulation meant that the action sequence to reach a goal was different across contexts but participants viewed the same visual stimuli. B) Example navigation trial in the scanner. Participants were first cued with a start and goal location and navigated through the maze one animal at a time. Inter-stimulus interval (ISI) was 3s. Arrows in red and blue indicate that participants had to make different actions to the same stimuli across contexts to reach their goal during navigation. C) Group level behavioral results (N = 23) from scanner showing elevated reaction times at decision points (Position 1 and Position 3). p1 > p2: $z = 13.97$, $p < 0.0001$; p1 > p3: $z = 9.13$, $p < 0.0001$; p1 > p4, $z = 11.67$ $p < 0.0001$; p3 > p2, $z = 4.84$, $p < 0.0001$; p3 > p4, $z = 2.536$, $p = 0.0112$; two-tailed, uncorrected. Pairwise comparisons were conducted using linear contrasts between estimated marginal means (z-test). Error bars represent +- SEM. * $p<0.05$, ** $p<0.001$. ITI = Interstimulus interval; SF = San Francisco; SD = San Diego

context. We next tested whether participants' reaction times would be modulated by differences in the decision-making demands at different locations in the virtual maze. Specifically, our task was structured such that participants were required to initiate their navigation plan at the onset of the start animal (i.e., position one), and at position three – the center of the plus maze, they needed to choose the correct move in order to reach the goal. Accordingly, we expected reaction times (RTs) to be higher at these positions in the navigational sequence than at other positions. Consistent with this prediction, analyses with a linear mixed effects model revealed a significant effect of position ($\chi^2(3, N = 23) = 220.99$, $p < 0.0001$), such that RTs were elevated at position one and position three, relative to other positions ($p1 > p2$: $z = 13.97$, $p < 0.0001$; $p1 > p3$: $z = 9.13$, $p < 0.0001$; $p1 > p4$, $z = 11.67$ $p < 0.0001$; $p3 > p2$, $z = 4.84$, $p < 0.0001$; $p3 > p4$, $z = 2.536$, $p = 0.0112$) (Fig. 1). This shows that decision-making demands at key locations, such as choice points, influenced participants' response time.

Hippocampus is sensitive to context-specific sequences in abstract spaces

During the planning phase (i.e., when participants were viewing the cues), we expected that participants should retrieve information about the sequence of state-action pairs that led from the start animal to the goal animal. Our first analyses targeted the extent to which hippocampal activity patterns carried information about the context and the planned sequence. To address this question, we extracted hippocampal multi-voxel activity patterns on each cue trial and calculated pattern similarity (Pearson's r) between trial pairs that came from repetitions of the same sequence cue in the same context, and compared those to both trial pairs for sequence cues with different start or

end points, and trial pairs for sequence cues that came from the same or different context (Fig. 2a). Importantly, visual information was shared across contexts as the cue only indicated the start and goal animal, not the context, and the same cue was associated with different moves between contexts. In addition, only trials which resulted in participants subsequently making the correct moves towards the goal were included in neural analyses.

To test whether hippocampal activity patterns carried information about the context and the planned sequence, we used a linear mixed effects model (Dimsdale-Zucker and Ranganath 2018) with fixed effects of context (same/different) and sequence (same/different), and a random intercept for subject (see Methods for model selection details and equation 2) to predict pattern similarity in the hippocampus. We reasoned that, during planning, participants retrieved information about the sequence of states and actions needed to reach the goal. Therefore, we predicted that pattern similarity should be higher for sequences that shared the same state-action pairs. Moreover, we predicted that this effect should be context-specific, as the same sequence across contexts have different state-action pairs. Consistent with this prediction, we found a significant sequence by context interaction (Fig. 2b: $\chi^2(1, N = 23) = 4.26$, $p = 0.04$). Follow up tests showed that patterns evoked by the same sequence cue in the same context were significantly different than all other trial pairs (same seq. + same cx. > diff. seq. + same cx.: $z = 2.77$, $p = 0.006$; same seq. + same cx. > same seq. + diff. cx.: $z = 2.73$, $p = 0.006$; same seq. + same cx. > diff. seq. + diff. cx.: $z = 2.61$, $p = 0.009$; see Fig. 2b). These results show that hippocampal activity patterns carried information about planned state-action sequences within specific contexts.

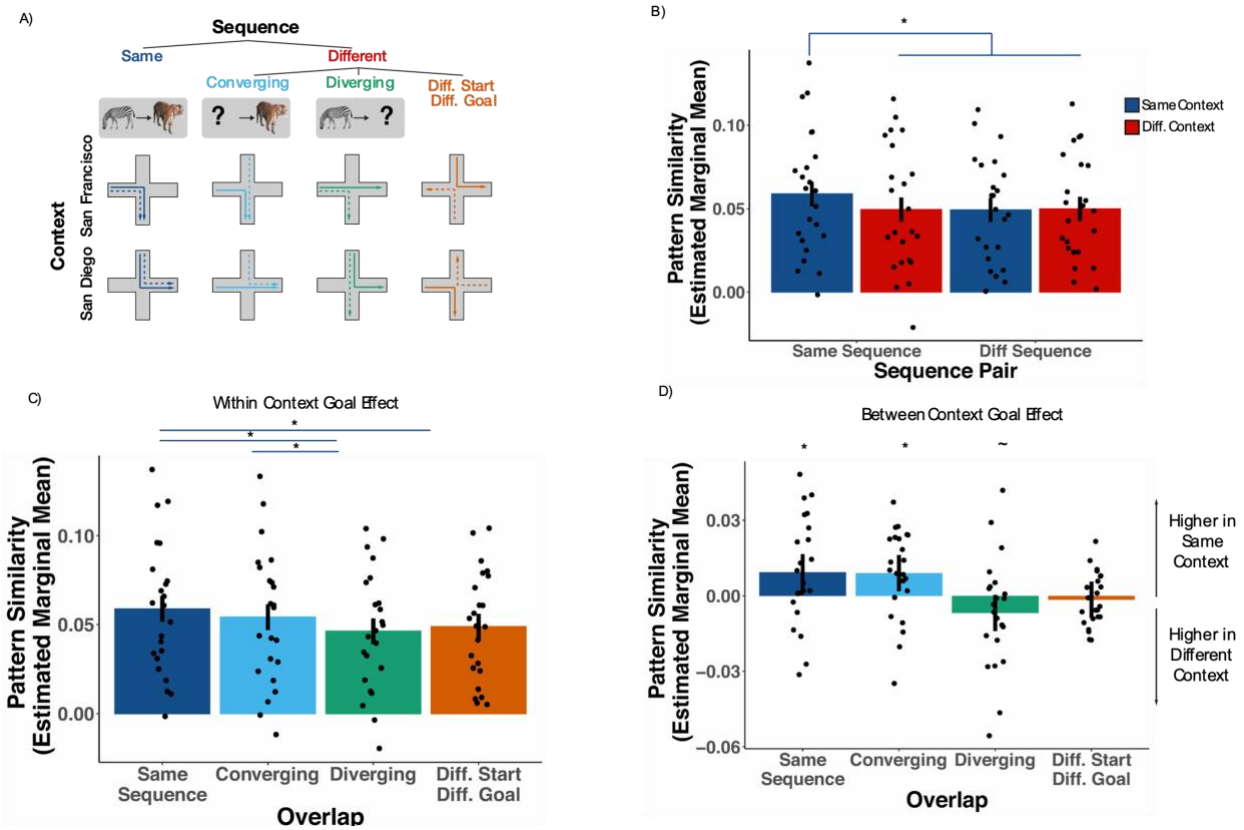

Fig. 2. Differential representation of future states in the hippocampus. A) Examples of trial pairs used in pattern similarity analyses during the planning phase. Dashed and solid lines of the same color represent two separate repetitions of the same trial type. B) Results from bilateral hippocampus. Pairs of trials sharing sequence and context have significantly higher pattern similarity than all other conditions (same cx. > diff. seq. + same cx.: $z = 2.77$, $p = 0.006$; same seq. + same cx. > same seq. + diff. cx.: $z = 2.73$, $p = 0.006$; same seq. + same cx. > diff. seq. + diff. cx.: $z = 2.61$, $p = 0.009$; two-tailed, uncorrected). C) Pattern similarity results comparing converging and diverging sequences within the same context. Same and converging sequences show higher similarity than diverging sequences (same seq. > diverging, $p = 0.0005$; same seq. > diff. start diff. goal, $p = 0.0056$; converging > diverging, $p = 0.03$; two-tailed, uncorrected). D) Pattern similarity results displaying the between context goal effect (interaction). Converging and same sequences show higher pattern similarity in the same context. Diverging sequences show higher pattern similarity in different contexts (same seq., $p = 0.0094$; converging, $p = 0.0012$; diverging, $p = 0.06$; diff. start diff. goal, $p = 0.67$; two-tailed tests, uncorrected). Pattern similarity was calculated using estimated marginal means obtained from linear mixed effects models. Pairwise comparisons were conducted using linear contrasts between estimated marginal means (z-test). Error bars represent 95% confidence intervals of the calculated estimated marginal means. Individual dots represent individual participants mean pattern similarity for each condition. N = 23. * $p < 0.05$, ~$p < 0.10$. cx = context; seq. = sequence.

Hippocampal activity patterns reflect future goals during planning

The above analysis demonstrates that hippocampal activity patterns carry context-specific information about planned sequences, but there are reasons to think

that hippocampal sequence representations might become more similar under certain circumstances. For instance, if the hippocampus uses predictive maps that carry information about possible future states (Stachenfeld et al., 2017), one might expect similar representations of “diverging” sequences that share the same starting point but lead to different goals by more heavily weighting the immediate state-action pairs that follow planning (see Methods for successor representation simulation details and Fig. S1). On the other hand, it is possible that goals are more heavily weighted during planning (Mattar and Daw, 2018), in which case we might expect similar representations of “converging” sequences that lead to the same goal but start at different states. We sought to test these ideas by comparing pattern similarity during cues associated with repetitions of the same sequence, cues associated with converging sequences that shared the same goal state, cues associated with diverging sequences that shared the same start state, and cues associated with sequences that had different start and different goal states (Diff. Start Diff. Goal)(Fig. 2a).

A linear mixed effects model with fixed effects for overlap (same sequence/converging/diverging/diff start + diff goal) and context (same/different) and a random intercept for subject (see Methods for model selection details and equation 3) showed a significant context by overlap interaction ($\chi^2(3, N = 23) = 14.75$, $p = 0.002$). (Fig. 2c and 2d**).** Follow up tests investigating this significant interaction revealed that, within a context, cues with converging goals had significantly higher pattern similarity than cues with diverging goals ($z = 2.19$, $p = 0.03$), and same sequence cues had higher pattern similarity than cues with diverging goals ($z = 3.49$, $p = 0.0005$) and cues with different starts and goals ($z = 2.77$, $p = 0.0056$). However, converging sequences

were not significantly different from the same sequence ($z = 1.30$, $p = 0.194$). Between contexts, cues of the same sequence and converging sequences showed significantly higher pattern similarity when in the same context (Same Sequence: $z = 2.60$, $p = 0.0094$; Converging: $z = 2.51$, $p = 0.012$). In contrast, diverging sequences showed a different pattern of results such that sequences from different contexts had higher similarity ($z = 1.89$, $p = 0.060$). Lastly, sequences with different starting states and goals were not significantly modulated by context ($z = 0.430$, $p = 0.67$). In sum, these results show that during planning, representations in the hippocampus are differentiated based on future context-specific goals. This suggests that goals may fundamentally shape representations in hippocampus via shared patterns between sequences that lead to the same goal.

Differences in pattern information during the cue period cannot be explained by shared motor plans or sensory details

The present results are consistent with the idea that the hippocampus supports planning of state-action sequences toward a goal. Importantly, our cues were carefully controlled, such that participants viewed visually identical stimuli across contexts and participants did not make responses during the planning phase. However, it is possible that low-level visual representations could be modulated by context (Huffman and Stark, 2017). To verify that visual regions did not show any effect of context, we ran a control analysis on an anatomically defined visual cortex ROI (V1/V2). To do this, we compared pattern similarity between cues of the same sequence, cues that had different starting items but the same goal, cues that had the same starting item but diverged to a different

goal, and cues that shared neither the start nor the goal. This analysis is identical to the overlap analysis run on hippocampus above (see Methods and equation 3 for model details). We found that this visual cortex ROI was only sensitive to visual information (Fig. S2 **-** Main effect of overlap – $\chi^2(3, N = 23) = 90.24$, $p < 0.001$ and not context ($\chi^2(1$ $N = 23) = 0.05$, Interaction: $p = 0.82$; $\chi^2(3, N = 23) = 0.76$, $p = 0.86$). This demonstrates that sensory representations of the cue were not modulated by context and likely do not drive any downstream contextual effects observed in the hippocampus.

Having verified that low-level visual information was not modulated by context, we next turned to representations of motor actions during panning. It is conceivable that, during planning, the pattern of results in hippocampus could be driven by overlap in planned movements between converging vs. diverging sequences. To ensure context effects observed in hippocampus were not due to shared motor information during planning, we examined trial pairs that had the exact same moves, trial pairs that had two moves in common, and pairs that had no moves in common to ensure that movement information alone was not modulated by context in the hippocampus. Results showed no effect of planned moves or context on pattern similarity (Fig S2 **-** main effect of context: $\chi^2(1, N = 23) = 0.46$, $p = 0.5$; main effect of move: $\chi^2(2, N = 23) = 1.56$, $p = 0.46$; interaction: $\chi^2(2, N = 23) = 2.68$, $p = 0.26$).

As a positive control analysis, we also examined an anatomically defined motor cortex ROI (BA4a/4p) to investigate whether we could detect sensorimotor representations and if they were modulated by context information during planning. Results revealed a significant main effect of planned move ($\chi^2(2, N = 23) = 13.95$, $p <$ 0.001), and importantly showed that planned movement was not modulated by context

(main effect: $\chi^2(1, N = 23), = 0.06$, $p = 0.81$; Interaction: $\chi^2(2, N = 23), = 0.68$, $p = 0.71$ Fig S2)(See Methods and equation 4 for model selection details). These results show that our cue period findings in the hippocampus cannot be solely explained by shared motor information of a plan and highlights the role of the hippocampus in retrieving the specific state-action sequence required to execute a plan. Altogether, these analyses provide an important control and bolster our interpretation of the findings from our analyses of the hippocampus, by showing that primary sensory areas are activating behaviorally-relevant representations during planning, but that the effects of context and goal are only present in hippocampus.

Representation of behaviorally relevant sequence positions during navigation

Having established that the hippocampus represents information about context-specific goals during planning, our next analyses turned to how state-action information is dynamically represented during navigation. Available evidence suggests at least three ways that navigationally-relevant information might be represented by the hippocampus. Based on classic studies of place cells, we might expect the hippocampus to represent the current state as participants navigated toward the goal. Alternatively, based on predictive map models (Stachenfeld et al., 2017), we could expect that the hippocampus would represent not only the current state but also future states.

A third possibility is that the hippocampus might preferentially represent goal-relevant information during navigation. In our study, the most behaviorally significant points in a navigated sequence were the starting point (position 1), when a goal-directed plan must be initiated, and the center of the maze (position 3), a critical sub-goal where

one's decision will determine the ultimate trial outcome. This was confirmed by our behavioral analyses that revealed that participants were slower to respond at positions 1 and 3 (Fig. 1). We therefore reasoned that participants might be likely to prospectively retrieve hippocampal representations of these states during navigation.

To test this prediction, we examined pattern similarity differences during navigation across converging and diverging sequences in the same zoo context. Converging and diverging sequences were chosen because these sequences have an equal number of overlapping states, but the timing of the overlap is systematically different. Both the "current state" and standard "predictive map" models would suggest that pattern similarity during navigation should reflect this pure overlap--early in a sequence there should be higher pattern similarity across pairs of diverging sequence trials, and late in a sequence there should be higher pattern similarity across pairs of converging sequence trials. In contrast, a goal-based account would predict that pattern similarity could reflect prospective coding of goal-relevant information (e.g. He et al., 2022, Brown et al., 2016) which should be higher across converging sequences (which share the same upcoming goal), relative to diverging sequences (which overlap in early states but lead to different goals).

We used a time-point by time-point pattern similarity analysis approach that enabled us to examine information in multivoxel activity patterns about current, past, and future states to test our key hypotheses. This technique is conceptually similar to cross-temporal generalization techniques used in pattern classification analyses (King and Dehaene, 2014). First, we extracted the time-series for each navigation sequence using a variant of single trial modeling that utilizes finite impulse response (FIR)

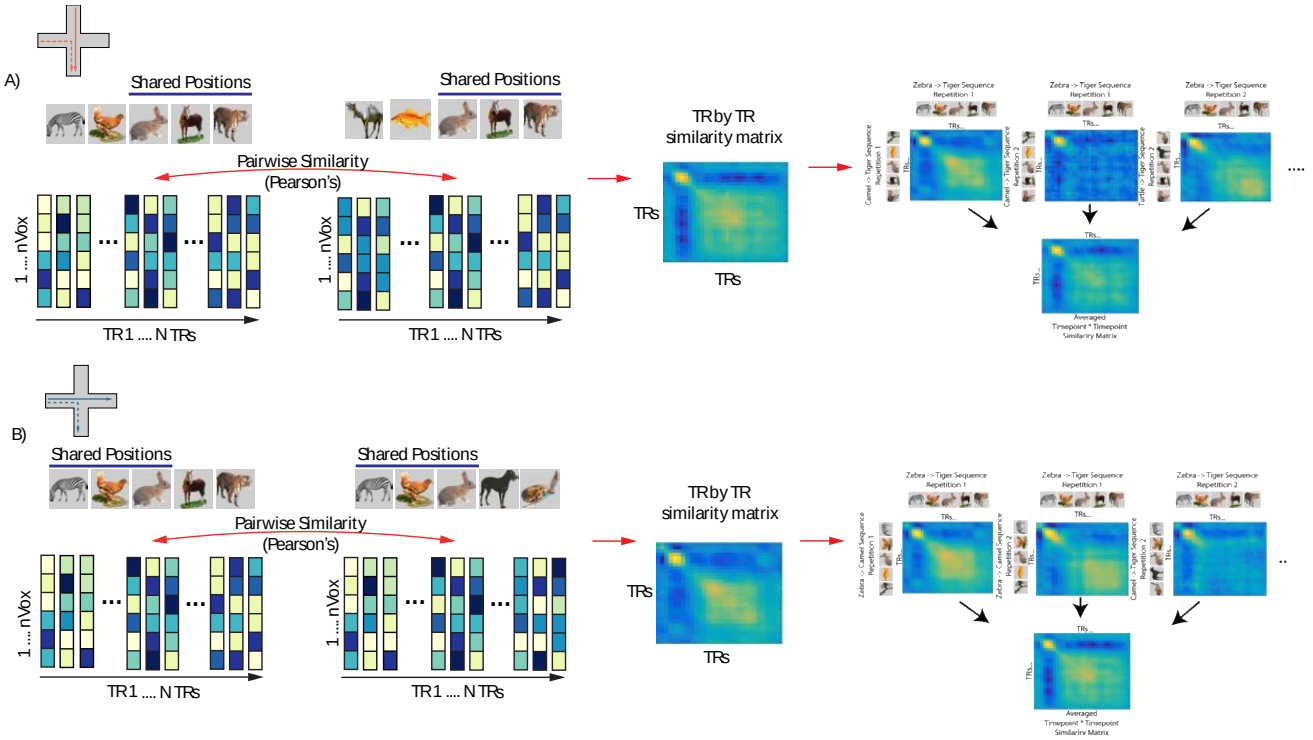

Fig 3. Schematic depiction of procedure to obtain time point by time point similarity matrices.
A) (Left) Dashed and solid lines on the maze indicate an example pair of trials correlated. TR by TR spatio-temporal patterns were obtained for a pair of sequences (converging in this example). Pattern similarity was computed between every possible pair of spatial patterns (voxels) over all timepoints (TRs) from a region of interest. (Middle) This procedure yielded a TR by TR similarity matrix for a given sequence pair. Note, that because the sequences are from different repetitions across fMRI scanning runs, the diagonal is not perfectly correlated. (Right) This was repeated for every possible converging sequence pair in the data set. The resultant TR by TR matrices were than averaged to create a subject level converging TR by TR matrix. Subject-specific averaged TR by TR matrices were than statistically compared to diverging sequences using cluster-based permutation tests (see Methods). B) Same as A but using an example diverging sequence pair.

functions (Turner et al., 2012), allowing us to examine activity patterns for each time point (TR) as participants navigated through the sequence of items. Importantly, incorrect trials were excluded from this analysis. As depicted in Figure 3, we quantified pattern similarity between pairs of navigation sequences (e.g. zebra to tiger sequence compared to camel to tiger sequence) at different timepoints (e.g., TR 1 to TR 10), which yielded a timepoint-by-timepoint similarity matrix for each condition (converging or diverging sequences). The diagonal elements for this matrix reflect similarity between pairs of animal items from the same timepoint in the sequence. Off-diagonal elements reflect the similarity between an animal at one timepoint in the sequence and animal

items at other timepoints in the sequence. Importantly, incorrect trials were excluded from this analysis.

Separate timepoint-by-timepoint correlation matrices (Pearson's r) were created for pairs of converging sequence trials and pairs of diverging sequence trials. We next computed a difference matrix and tested for statistically significant differences between converging and diverging sequences, correcting for multiple comparisons using cluster-based permutation tests (10,000 permutations, see Methods for more details).

As noted above, diverging sequences have overlapping states early in the sequence, and converging sequences have overlapping states late in the sequence. If the hippocampus represents only current states, we would expect to see pattern similarity differences between converging and diverging close to the diagonal of the timepoint-by-timepoint matrices — that is, we would expect higher pattern similarity for diverging pairs during timepoints early in the sequence and higher pattern similarity for converging pairs during timepoints late in the sequence. If the hippocampus represents current and temporally-contiguous states, as suggested by predictive map models, we would expect that at early positions, we would expect higher pattern similarity for diverging sequences, both on- and off –diagonal, and at late positions, we would expect higher pattern similarity for converging sequences both on- and off –diagonal. Finally, if the hippocampus preferentially represents goal-relevant information during navigation (Mattar and Daw, 2018, He et al., 2022), we would expect to see higher off-diagonal pattern similarity only for *converging* sequences, because only converging sequences share the same goal. Specifically, we expected higher off-diagonal pattern similarity between goal states and earlier positions in the sequences.

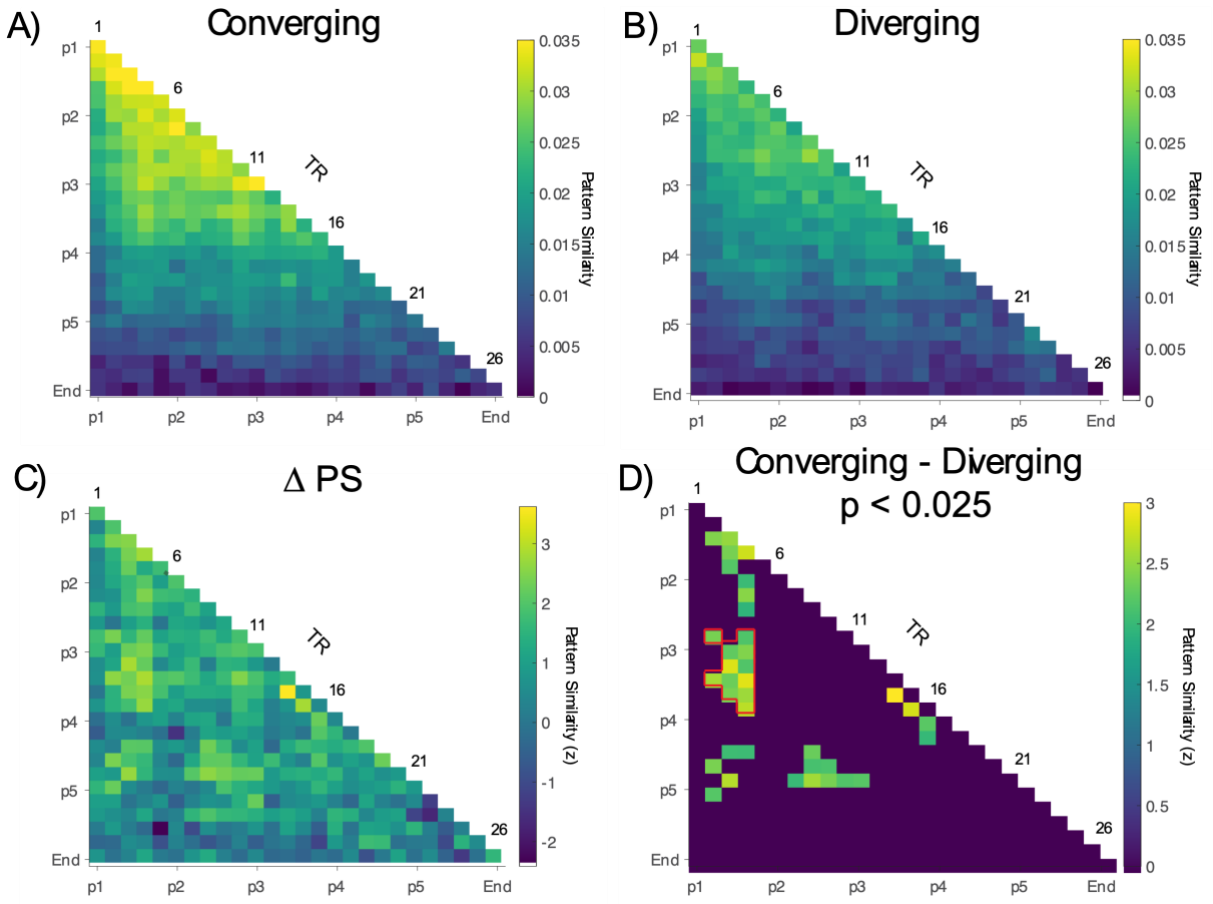

Fig. 4. Results from TR by TR pattern similarity analysis during active navigation in bilateral hippocampus. A) Group level pattern similarity results from converging sequences during active navigation. B) Same as A but showing diverging sequences. C) TR by TR pattern similarity results depicting a statistical map of converging – diverging. Z values were calculated using a bootstrap shuffling procedure with 10,000 permutations. D) Thresholded statistical map at $p < 0.025$ (two-tailed). Cluster based permutation tests with 10,000 permutations (Maris and Oostenveld, 2007) were performed with a cluster defining threshold of $p < 0.025$ (two-tailed) and a cluster alpha of 0.05 (two-tailed). Outlined in red is a significant cluster of timepoints that survives multiple comparisons correction (cluster mass = 29.44, $p = 0.038$, maximum cluster corrected). Note that this cluster corresponds to approximately position 1 activating position 3 which was shared by both converging and diverging sequences. Trial labels were manually lagged by 4 TRs (TR = 1.22, Inter-Item-Interval = 5s) to account for hemodynamic response lag. In panels C and D, each pixel of a statistical comparison (T-value, N = 23) was converted into a Z value by normalizing it to the mean and standard error generated from our permutation distributions (see Methods).

Consistent with the prospective representation of goal-relevant states in the hippocampus, we found several clusters showing higher similarity for converging compared to diverging sequences (Fig. 4). Interestingly, there was a significant off-diagonal cluster (outlined in red: $p = 0.038$, corrected) that roughly corresponds to the activation of the decision point (position 3) when participants were at position 1 (approx. TRs 10-15). Other clusters tended to overlap with key locations in the experiment, which

roughly correspond to position one activating position five (TRs 18 to 21) and position three activating position five (TRs 18 to 20) (Fig. 4e), although these clusters did not survive multiple comparison correction. These data are consistent with the idea that information about position 3 was preferentially activated in converging sequences, in which the same key decision was required to navigate to the same goal.

Discussion

The aim of the present study was to identify how the hippocampus represents task information during planning and navigation towards a behavioral goal. During planning, we show that hippocampal representations carried context-specific information about individual sequences to a goal. Surprisingly, not all sequences were equally differentiated, such that sequences that converged on a common goal showed higher pattern similarity compared to diverging sequences, despite an equal amount of overlap between the conditions. Similarly, during navigation, we found that the hippocampus prospectively activated goal-specific representations of the key decision point. Taken together, our results suggest that the hippocampus forms integrated representations of sequences that lead to the same goal. Furthermore, they support the notion that the hippocampus plays a phasic role in the activation of patterns that contain information about future states and prioritizes sub-goal information during active navigation. In summary, our data are consistent with the idea that rather than simply representing overlapping associations, hippocampal representations are powerfully shaped by context and goals.

The hippocampus represents context-specific goal information during planning

A key finding from the present study is that, during planning, hippocampal activity patterns are organized such that they either generalize or differentiate between sequences depending on the goal, and do so in a context-specific manner. These findings are relevant to theories which propose that prospective thought (prediction/planning) relies on the same circuitry used for episodic memory (Hassabis et al., 2007; Schacter et al., 2007, Addis et al., 2012). In support of this idea, place cells fire in a sequence that represents the path that an animal will take in a phenomenon described as “forward replay” (Johnson and Redish, 2007, Pfieffer and Foster 2013). This work supports the hypothesis that look-ahead processes at the single neuron level may support planning (but see Gillespie et al., 2021). Building on this work, Brown et al., (2016) used high-resolution fMRI in humans to examine hippocampal activity during goal-directed navigation in a virtual reality (VR) paradigm. Brown et al. demonstrated that, during planning, hippocampal activity patterns could be used to accurately decode future navigation goals, even across different start positions and routes. Thus, their findings demonstrated that fMRI activity patterns in the hippocampus carried information about future navigational goals. Brown et al. (2016) interpreted their findings as evidence that the hippocampus supports imagination or mental simulation of a route towards a goal.

Our findings suggest an important constraint on the role of the hippocampus in imagination and simulation. In our study, if participants simulated the sequence of sensory events that led to the goal (i.e., imagining the sequence of animals), we would expect hippocampal representations to generalize across repetitions of the same

sequence of animals across the two different zoo contexts. Instead, we found that hippocampal representations during planning were context specific, such that pairs of trials involving the same sequence of animals across different contexts were indistinguishable from entirely different sequences. Moreover, similarity across different sequences that led to the same goal in the same zoo context was indistinguishable from similarity across repetitions of the same sequence in the same context. Thus, in our study, hippocampal activity most likely did not reflect imagination of a sequence of stimuli per se, or even a specific sequence of states, but rather a context-specific representation of behaviorally relevant points to achieve a goal.

Together with prior research, our results are relevant to an emerging body of work suggesting that goals and other salient locations exert a powerful force on spatial and non-spatial maps in the brain (McKenzie et al., 2013, 2014; Boccara et al., 2019; Butler and Hardcastle et al., 2019; Brunec et al., 2018). For example, McKenzie et al., (2014) found that rewarded events had higher pattern similarity within a context compared to unrewarded events. Moreover, there is evidence that, after learning in a reward-based foraging task, place cells tend to be clustered around goal locations (Dupret et al., 2010, Gauthier et al., 2018). This could go some way towards explaining our results of increased pattern similarity for sequences that converge on the same goal. Considering the current work and past findings, we propose that hippocampal representations are flexibly modulated depending on current behavioral demands, incorporating trial-specific information that allows organisms to realize a specific goal (Ekstrom and Ranganath, 2017).

Our findings are also relevant to past work showing that the hippocampus represents information about specific sequences of objects (Hsieh et al., 2014; Schapiro et al. 2016; Kalm, Davis, and Norris, 2013; Agster, Fortin and Eichenbaum 2002; Bellmund et al., 2022; Allen et al. 2016). Studies examining how the brain represents routes with multiple paths or that are hierarchical in nature show that activity in the hippocampus is higher when planning and navigating an overlapping route and that, during navigation, univariate bold signal is modulated by distance to a goal (Brown et al., 2014; Balaguer et al., 2016; Chanales et al., 2017). In one study, Chanales et al. (2017) show that representations of overlapping spatial routes become dissimilar over learning. This is potentially at odds with the current findings, where we find that routes that overlap in their goal show higher pattern similarity compared to routes that do not share a goal. However, participants in Chanales et al. (2017) passively viewed pictures along routes, whereas participants in our task actively navigated. As mentioned earlier, rodent studies suggest that hippocampal spatial coding can shift dramatically between goal-directed behavior and random foraging in the same context. Moreover, in Chanales et al. (2017) it would make sense for participants to differentiate overlapping routes because they did not include sequences that converged on the same goal. Thus, it would be optimal to learn a unique representation for each spatial route in order to predict the outcome. In contrast, in our experiment, all trials that converged on the same goal required the same key decision at position 3, regardless of the starting point. In this situation, it is optimal to learn a representation that captures the information that is common to any sequence that converges on the same goal. For example, as depicted in Figure 1, any trial with a tiger as the goal animal will require participants to choose

the “down” button at position 3. In the next section, we explain why results from the navigation period are also consistent with this interpretation.

Reinstatement of remote timepoints in the hippocampus during navigation

If the hippocampus supports prospective planning for goal-directed navigation, then it is important to understand how it functions when such actions are taken when navigating abstract spaces. For example, if the hippocampus is involved in retrieving the specific state-action plan, what is its function once this plan is executed? To address this question, we contrasted pattern similarity during the navigation phase across pairs of converging sequences against pairs of diverging sequences.

As noted above, the animals in the first three positions overlapped across diverging sequences, whereas the animals in the last three positions overlapped across converging sequences. Thus, if the hippocampus only represented the current state during navigation, we would have expected pattern similarity on the diagonal in Figure 4 to be higher for diverging trials for early time points, and then higher for converging trials in the later time points (see also Figure S4). If participants solely retrieved past states during navigation, we would expect off-diagonal pattern similarity to be higher for diverging sequences than converging sequences (because the first three positions were common for the diverging sequences). Our data were inconsistent with both of these accounts. Instead, we found that off-diagonal pattern similarity was higher for converging than for diverging trial pairs, suggesting that hippocampal activity patterns carried information about future timepoints during navigation.

The significant cluster of increased pattern similarity for converging, relative to diverging, sequences was consistent with the interpretation that, at the outset of the navigation phase, participants prospectively activated a representation of position 3. This result is notable for two reasons. First, participants were engaged in active, self-initiated navigation, and as such, we would expect considerable variability in the timing of prospective coding across trials and across subjects. The fact that prospective coding of position 3 (as indicated by off-diagonal pattern similarity) was nonetheless reliable across participants attests to the significance of this position to successful task performance. Second, the finding is notable because the stimulus at position 3 is exactly the same for all trials in all contexts. Thus, the disproportionate representation of position 3 across convergent sequences could not solely reflect the identity of the stimulus itself.

As noted above, the correct decision to be made at position 3 depends on one's current goal and context. All converging sequences share the same decision at position 3 because they share the same goal, whereas diverging sequences are associated with different decisions at position 3 because they involve different goal states. These results are consistent with the idea that participants prospectively activated the most goal-relevant information in the upcoming sequence, namely the context- and goal-appropriate decision at position 3.

Consistent with our study, research in rodents shows that hippocampal ensemble activity differs between routes that share a common path but lead to a different goal (Frank et al., 2000; Wood et al., 2000, Ferbinteanu and Schapiro, 2003; Ito et al., 2015; Markus et al., 1995). There are also findings that demonstrate predictive hippocampal

representations that are related to future behavior in both spatial and non-spatial tasks (e.g. Garvert et al., 2017, Brunec and Momennejad, 2022). Our data, however, suggest that, during goal-directed behavior, the human hippocampus does not solely reflect the current state during navigation, or only the immediate future, but rather that it emphasizes strategically important states for reinstatement during ongoing behavior. Our results align with computational models that show that place cells associated with behaviorally relevant locations in an environment are preferentially incorporated into replay events (e.g. Mattar and Daw, 2018).

Relevance to models of hippocampal state space representation

Several models of hippocampal contributions to spatial navigation and memory propose that the hippocampus generates predictions of upcoming states (e.g. Barron et al., 2020). For instance, a specific computational implementation of a predictive map model, the successor representation, states that the hippocampus is involved in learning relationships between states and actions, and that its representations reflect expectations about future locations (Stachenfeld et al., 2017; Momennejad, 2020). We used a standard version of this computational model to generate simulated pattern similarity results, and surprisingly, these simulated matrices were qualitatively different from what we observed in the hippocampus.

In our simulations (see Supplemental Materials), a classical version of the successor representation reflected the transition probabilities between states, such that adjacent states were more similar than non-adjacent states. Because participants transitioned between all start and end positions equally in both directions, the model

could not reproduce the difference between converging and diverging sequences either during the planning or navigation phases. It is possible that, in the relatively small and deterministic state space used in our task, it is not advantageous to represent an elaborate transition structure. An alternative approach to account for the present results would be to use a model that places heavier emphasis on context instead of only the next item or next decision. One model, the “clone-structured cognitive graph” model (George et al., 2021), is able to learn “clones” of similar observations that are distinguished by the current context. We predict that that models that take into account context and goals, like the model presented in George et al., will be better able to capture the nuances of our task.

Alternatively, it might be advantageous to focus on models that incorporate an inductive bias to specifically focus on the most goal-relevant aspects of a state space (e.g., the goal, context, and decision at P3). In many situations, an agent with an appropriate understanding of task structure could benefit by deploying the hippocampus more strategically, by preferentially encoding and prospectively retrieving memories for key decision points towards a goal (O’Reilly, Russin, & Ranganath, 2022). One example of a computational model that relies on strategic deployment of past experience comes from Lu, Hasson, and Norman (2022). Their simulations showed that it was computationally advantageous to prioritize hippocampal encoding and retrieval of temporally extended events at event boundaries, which are moments of high uncertainty or prediction error. In our task, inductive biases carried out through such a computational framework could emphasize the goal and key decision point, rather than passively representing all possible state transitions.

We believe that hippocampal representations of physical space (Ekstrom and Ranganath, 2017) and abstract state spaces (Boorman, Sweigert, and Park, 2021) are flexible, reflecting the computational demands of the planning problem, and the subject's understanding of, and experience with, the problem. In the present study, the task might have encouraged a model-based planning strategy in which future goals and key states are strategically retrieved and represented in hippocampus. In cases where learning is passive and incidental to the task, or when transitions between states change unpredictably, hippocampal state spaces might instead resemble successor-based maps. Finally, in more complex tasks, participants might adopt different strategies with varying degrees of emphasis on goal-relevant information (See Eldar et al., 2020).

Human behavior is characterized by the ability to plan and flexibly navigate decision spaces in order to realize future goals. The present findings suggest that the hippocampus represents context-specific, goal-oriented representations during navigation. These findings may contribute to the development of unified models accounting for hippocampal contributions to memory, navigation, and goal-directed sequential decision-making (Eichenbaum, 2017; Wikenheiser and Schoenbaum 2016; Bellmund et al., 2018). Additionally, this work highlights the importance of studying goal-directed behavior, attentional modulation of memory representations, and their impact on planning.

Methods

Participants: Thirty healthy English-speaking individuals participated in the fMRI study. All participants had normal or corrected-to-normal vision. Written informed consent was obtained from each subject before the experiment, and the Institutional Review Board at the University of California, Davis approved the study. Participants were compensated with an Amazon gift card for their time. Data from one participant was excluded due to technical complications with the fMRI scanner, one participant was excluded due to a stimulus computer malfunction, two participants were excluded due to poor behavioral performance in the scanner (defined as falling below trained criterion, 85% correct, in the scanner), and one participant was removed from the scanner before the experiment concluded because they did not wish to continue in the study. Prior to data analysis, to ensure data quality, we conducted a univariate analysis to examine motor and visual activation during the task compared to an implicit baseline (unmodeled timepoints when the participant was viewing a fixation cross). Two participants showed little to no activation in these regions and were excluded from further analysis. The remaining 23 participants (11 male, 12 female, all right handed) are reported here.

Stimuli and Procedure: Data was collected from participants using Matlab 2016a and Psychophysics toolbox. Task stimuli consisted of nine common animals, shown in color on a grey background. Participants were tasked with learning two “zoo contexts”, consisting of animals organized in a specific spatial orientation (Figure 1a). Importantly, animals in both contexts were visually identical, but each context had a distinct spatial

organization. Training consisted of three stages per context: 1) map study, 2) exploration, 3) sequence navigation. This was followed by an additional sequence navigation phase that alternated between contexts.

During map study, participants were initially shown an overhead view of one of the zoo contexts (counterbalanced across participants). After studying this picture, participants were asked to reconstruct the location of all the animals by arranging icons on the screen. If participants were not able to perfectly recreate the maze they were shown the picture once more and asked to try again. Next, during the zoo exploration, participants used arrow keys to move between items in the zoo, starting from the central animal. At the bottom of the screen participants were shown arrows indicating all possible moves from their current location (e.g. Left, Up, Down, Right at the center position of a maze). If participants made an incorrect move (moving outside of the animal maze) they were informed they made a wrong move. Participants were required to visit each of the animals four times before moving on to the next phase. During the sequence navigation phase, participants were shown a cue with a start and goal animal, and had four moves to reach the goal on a given trial. Start and goal animals were always the endpoints of an arm. Participants were trained to 85% criterion before learning the other context. The same training procedure outlined above was repeated for the second zoo context. After learning each of the zoos to criterion, participants completed an additional sequence navigation phase with the same timing as the MRI scanning session.

In the MRI scanner, participants completed six runs of the sequence navigation task (Figure 1b). In each run, participants completed 16 sequence navigation trials. Trials from a given context were presented in blocked fashion so that there were 8 consecutive trials from each context. Across runs, context blocks were alternated and their order was counterbalanced across participants. In addition, the order of sequences within each context was counter-balanced across blocks to ensure no systematic ordering effects influenced our results. Each navigation trial began with a cue signaling a start and a goal animal displayed for 3s, followed by a 3s ITI. Participants then saw the start animal and navigated by pressing buttons to move through the space one animal at a time. Animal items were displayed on the screen for 2s with a 3s ITI, regardless of participant button press. For items where participants made a navigational error, text was displayed for 2s informing them they made a wrong move or incorrectly navigated to a goal animal. In each zoo context, participants planned and navigated 12 distinct sequences (each repeated 4 times across 6 runs of scanning)

MRI Data acquisition: MRI data were acquired on a 3T Siemens Skyra MRI using a 32-channel head coil. Anatomical images were collected using a T1-weighted magnetization prepared rapid acquisition gradient echo (MP-RAGE) pulse sequence image (FOV = 256 mm; TR = 1800 ms; TE = 2.96 ms; image matrix = 256 x 256; 208 axial slices; voxel size = 1mm isotropic). Functional images were collected with a multi-band gradient echo planar imaging sequence (TR = 1222 ms; TE = 24 ms; flip angle = 67 degrees; matrix=64x64, FOV=192mm; multi-band factor = 2; 3 mm^3 isotropic spatial resolution).

MRI data processing: Data were preprocessed using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/) and ART Repair (Mazaika et al., 2009). Slice timing correction was performed as implemented in SPM12. We used the iterative SPM12 functional-image realignment to estimate movement parameters (3 for translation and 3 for rotation). Motion correction was conducted by aligning the first image of each run to the first run of the first session. Then all images within a session were aligned to the first image in a run. No participant exceeded 3mm frame wise displacement. A spike detection algorithm was implemented to identify volumes with fast motion using ART repair (0.5mm threshold) (Power et al., 2012). These spike events were later used as nuisance variables within generalized linear models (GLMs). Participants native structural images were coregistered to the EPIs after motion correction. The structural images were bias corrected and segmented into gray matter, white matter, and CSF as implemented in SPM12. Native brainmasks were created by combining gray, white matter masks. Data were smoothed with a 4 mm^3 FWHM 3D gaussian kernel.

Regions of Interest: ROI definitions were generated using a combination of Freesurfer, and a multistudy group template of the medial temporal lobe. The multistudy group template was used to generate probabilistic maps of hippocampal head, body, and tail as defined by Yushkevich et al. (2015), and warped to MNI space using Diffeomorphic Anatomical Registration Using Exponentiated Lie Algebra (DARTEL) in SPM8. Maps were created by taking the average of 55 manually-segmented ROIs and therefore reflect the likelihood that a given voxel was labeled in a participant. Masks were created by thresholding the maps at 0.5, (i.e., that voxel was labeled in 50% of participants). These maps were then reverse normalized to native subject space using Advanced

Normalization Tools (ANTS). Subject specific cortical ROIs were generated using Freesurfer version 6.0. from the Destrieux and Desikan atlas (Desikan et al., 2006, Fischl et al., 2004, Desitrieux et al., 2010). Individual cortical ROIs were binarized and aligned to participants' native space by applying the affine transformation parameters obtained during coregistration. These masks were combined into merged masks that encompassed the entire hippocampus bilaterally (see cue period pattern similarity for more information). Anatomical ROIs for V1/V2 and BA4a/p were obtained by running all participants structural scans through the freesurfer recon-all pipeline. Our V1/V2 ROI was obtained by merging the anatomical masks for BA17 and BA18 **(**Fig. S2**).**

Cue period pattern similarity analysis: Our primary interest was to investigate how prospective sequence representations were modulated based on context membership. To achieve this, we used representational similarity analysis to analyze multi-voxel activity patterns (Kriegeskorte et al., 2008) within regions of interest. Generalized Linear Models (GLMs) were used to obtain single trial parameter estimates of the cue period using a modified least-squares all (LSA) model (Mumford et al., 2012, Brown et al., 2016). Data were high-pass filtered using a 128s cutoff and pre-whitened using AR(1) in SPM. All events were convolved with a canonical HRF to be consistent with prior work (Mumford et al., 2012). Cue periods were modeled using separate single trial regressors for each cue (2s boxcar). The remaining portions of the task were modelled as follows: Navigation periods were modelled with separate 25s boxcar functions for each trial, separate single trial regressors for catch sequences modelled as a 15s boxcar, separate single trial catch blank trials (stick function), outcome correct at condition level (stick), outcome incorrect at condition level (stick), and the four button presses at the

condition level (stick). Nuisance regressors for motion spikes, 12 motion regressors (6 for realignment and 6 for the derivatives of each of the realignment parameters) and a drift term were included in the GLM. Pattern similarity between the resulting beta images were calculated using Pearson's correlation coefficient between all pairs of trials in the experiment. Only between run trial pairs were included in the analysis to avoid spurious correlations driven by auto-correlated noise (Mumford et al., 2014).

Based on evidence of functional differentiation along the long-axis of the hippocampus (Poppenk et al., 2013, Bouffard et al., 2021, Brunec and Momennejad., 2022), we tested for any longitudinal or hemispheric differences in hippocampal patterns. Analyses revealed no significant differences in the pattern of results between left and right or between anterior or posterior segments of the hippocampus. As a result, subsequent analyses were performed with pattern similarity data from a bilateral hippocampus mask.

Linear mixed models: Behavioral responses and pattern similarity were analyzed using linear mixed effects models to account for the nested structure of the dataset, allowing us to statistically model errors in our model clustered around individuals and trial types that violate the assumptions of standard multiple regression models. Statistical comparisons were conducted in R (3.6.0) (https://www.r-project.org/) using lme4 (Bates et al., 2015) and AFEX (Singman et al., 2016). Reaction times were analyzed using the following formula:

(1) (Figure 1): RT ~ Position + (1|subject)

Where (1|x) indicates the random intercept for subject and RT is the reaction time for each position during the navigation phase, excluding position 5 (as no response is required). Furthermore, outlier RTs were excluded that exceeded 2.5 standard deviations from a participants average reaction time.

For the pattern similarity analyses, pairwise PS values were input for each subject into three separate models with the following formulas:

(2) (Figure 2b): PS ~ same_sequence*same_context + (1|subject)

(3) (Figure 2c/d): PS ~ overlap*same_context + (1|subject)

(4) (Figure S2): PS ~ move*same_context + (1|subject)

Where (1|x) indicates the random intercept for subject and PS is the Pearson correlation coefficient for a given trial pair. Fixed effects for equation 1: (1) same sequence - a categorical variable with two levels indicating if the trial pair was from the same or different sequence. (2) Same context - categorical variable with two levels: same or different. Fixed effects for equation 3: overlap - a categorical variable with four levels: full, converging, diverging, and diff. start diff. goal. Same context - same as equation 2. Fixed effects for equation 4: Move - a categorical variable with three levels: same moves, shared moves, no moves. Same context - same as equation 2. Statistical significance for fixed effects was calculated by using likelihood ratio tests, a non-parametric statistical test where a full model is compared to a null model with the effect of interest removed. For example, to test the significance of an interaction term two models would be fit. One with two main effect and no interaction and the other with the

interaction term. Follow up tests and estimated marginal means (Searle et al., 1980) from LMMs were calculated using the R package emmeans (https://cran.rproject.org/web/packages/emmeans/index.html).

In all the above models, a model with a maximal random effects structure, as recommended by Barr et al., 2014, was first fit. In all cases the maximal model failed to converge or was singular, indicating over-fitting of the data. When examining the random effects structure for these models, random slopes for our fixed effects accounted for very little variance when compared to our random intercept for subject. To improve our sensitivity and avoid over-fitting these terms were removed as suggested by Matuschek et al., 2017. Lastly, it is important to note that our results are not dependent on using linear mixed models. Using a standard repeated measures ANOVA produces qualitatively and quantitatively similar results in all ROIs.

Successor Representation Simulation: To better understand specific predictions of the successor representation in our task (Stachenfeld et al., 2017) we performed a simple simulation with respect to our task. First, we created a topological structure (connected graph) that was similar to our task. As seen in Figure S1, this structure closely resembled the plus maze participants navigated in. We simulated the successor representation based on a random walk policy using the equation.

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

Where γ is a free parameter that controls the decay of the SR and T is the full transition matrix of the task depicted in Fig. S1A/B. For the current simulations, gamma of 0.3 was

used, but results are qualitatively similar for different values. Random walk or policy independence can be assumed in this case because maps were well learned before the scanner and each sequence was traversed in both directions an equal number of times (Momennejad, 2020).

We then tested the hypothesis that, during planning, the hippocampus encodes the SR of the first position in the sequence (columns of SR). We extracted columns of the SR for three planned sequences ((state 1 -> state 5) (state 6 -> state 5) (state 1 -> 9)) and calculated the similarity (Pearson's) between pairs of trials. The same sequence was calculated by correlating the same sequence with itself. The converging condition was obtained by correlating trials that started at different states but converged on the same end state. The diverging condition was obtained by correlating trials that started at the same state but diverged to different end state. Lastly, the diff. start diff. goal condition was calculated by correlating trial pairs that started and ended at different states. As shown in Figure S1, the SR heavily weights the immediate locations around the starting location and thus would predict that diverging sequences should have higher similarity than converging sequences.

Timepoint-by-timepoint representational similarity analysis:

To examine whether participants activated remote timepoints as they navigated through our virtual environments (e.g., activating decision points early in the navigation trial), we used a variant of single trial modeling using finite impulse response (FIR) functions (Turner et al., 2012). This method allowed us to isolate the unique spatiotemporal pattern of activity for a given navigation trial while simultaneously controlling for

surrounding time points during the run. We modeled 47 seconds of neural activity with a set of 38 FIR basis functions. Specifically, we obtained a spatial pattern of activity for each of these 38 TRs in our model, which allowed us to compare the similarity of the spatial patterns of activity between timepoints in the navigation phase. Additional regressors were included for motion, however spike regressors were not included in this analysis because they perfectly colinear with an FIR basis sets for each TR. A separate GLM was used for every trial resulting in 72 voxel time series. Collinearity in our model was measured using the variance inflation factor (VIF) and was verified to be within acceptable levels according to standards in the literature (Mumford et al., 2015). To examine within trial type similarity (same trial type across repetitions) timepoint-by-timepoint similarity matrices were generated by correlating activity patterns from repetitions of specific sequence pairs (e.g. zebra-tiger repetition 1 with camel-tiger repetition 1), at every TR. The resultant matrices were symmetrized by averaging across the diagonal of the matrix using the following equation: $(X^T + X)/2$. The resultant timepoint-by-timepoint similarity matrix was averaged within a specific trial type to get a single average timepoint by timepoint similarity matrix for each subject and condition (Fig. 3). This was done separately for converging and diverging sequences. Only between run trial pairs were included in the analysis to avoid spurious correlations driven by auto-correlated noise (Mumford et al., 2014). This method allowed us to isolate individual sequence patterns while controlling for temporally adjacent navigation trials. To identify which points in time corresponded to relevant parts of the task, we manually lagged trial labels by 4 TRs to account for the slow speed of the HRF.

Time point by timepoint similarity matrices were constructed only for converging and diverging sequences. This subset of trials was chosen for several methodological reasons listed below. One is that, to maximally control for differences in trial numbers between conditions and temporally auto-correlated evoked patterns, while still maintaining enough power to examine future state reactivation; we restricted our analyses to converging and diverging sequences within the same context. Importantly, this selection of trials allows us to simultaneously control for several factors while testing specific predictions. Another is that, converging and diverging sequences are matched in terms of the number of shared items and therefore overall visual similarity. Specifically, the same animal items are seen during the first half of diverging sequences, while the same animal items are seen in the second half of converging sequences (all sequences share the center item).

To assess statistical significance, and to correct for multiple comparisons, we used cluster-based permutation tests (Marris and Oostenveld, 2007) with 10,000 permutations, with a cluster defining threshold of 0.05 (two-tailed). Each pixel of a statistical comparison (T-value) was converted into a Z value by normalizing it to the mean and standard error generated from our permutation distributions. Cluster significance was determined by comparing the empirical cluster size to the distribution of the maximum cluster size (sum of T-values) across permutations with a cluster mass threshold of 0.05 (two-tailed).

Data Availability: Processed data to reproduce figures in the manuscript and supplement are available at: https://github.com/jecd/Hippocampgoal. Source data are provided with this paper. A reporting summary for this article is available as a supplementary information file.

Code availability: Code to reproduce all figures and statistical analyses in the manuscript and supplement are available at: https://github.com/jecd/Hippocampgoal

References

Addis, D. R., & Schacter, D. L. (2012). The hippocampus and imagining the future: Where do we stand? *Frontiers in Human Neuroscience*, *5*(JANUARY 2012), 1–15. https://doi.org/10.3389/fnhum.2011.00173

Agster, K. L., Fortin, N. J., & Eichenbaum, H. (2002). The hippocampus and disambiguation of overlapping sequences. *Journal of Neuroscience*, *22*(13), 5760–5768. https://doi.org/10.1523/jneurosci.22-13-05760.2002

Ainge, J. A., Tamosiunaite, M., Woergoetter, F., & Dudchenko, P. A. (2007). Hippocampal CA1 place cells encode intended destination on a maze with multiple choice points. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(36), 9769–9779. https://doi.org/10.1523/JNEUROSCI.2011-07.2007

Allen, T. A., Salz, D. M., McKenzie, S., & Fortin, N. J. (2016). Nonspatial Sequence Coding in CA1 Neurons. *Journal of Neuroscience*, *36*(5), 1547–1563. https://doi.org/10.1523/JNEUROSCI.2874-15.2016

Alme, C. B., Miao, C., Jezek, K., Treves, A., Moser, E. I., & Moser, M. B. (2014). Place cells in the hippocampus: Eleven maps for eleven rooms. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18428–18435. https://doi.org/10.1073/pnas.1421056111

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, *543*(7647), 719–722. https://doi.org/10.1038/nature21692

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

Barr, Dale., Levy Roger., Scheepers, Christoph., Tily, H. (2014). Random effects structure for confirmatory hypothesis testing: Keep it maximal, *68*(3), 1–43. https://doi.org/10.1016/j.jml.2012.11.001.Random

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology*, *192*(October 2019). https://doi.org/10.1016/j.pneurobio.2020.101821

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

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

Bellmund, J. L. S., Deuker, L., Montijn, N. D., & Doeller, C. F. (2022). Mnemonic construction and representation of temporal structure in the hippocampal formation. *Nature Communications*, *13*(1), 1–16. https://doi.org/10.1038/s41467-022-30984-3

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

Boccara, C., Nardin, M., Stella, F., O'Neill, J., & Csiesvari, J. (2019). The entorhinal cognitive map is attracted to goals. *Science (New York, N.Y.)*, *363*(March), 1443–1447. https://doi.org/10.1126/science.aav4837

Boorman, E. D., Sweigart, S. C., & Park, S. A. (2021). Cognitive maps and novel inferences: a flexibility hierarchy. *Current Opinion in Behavioral Sciences*, *38*, 141–149. https://doi.org/10.1016/j.cobeha.2021.02.017

Bouffard, N., Golestani, A., Brunec, I. K., Bellana, B., Barense, M., & Moscovitch, M. (2021). Single voxel autocorrelation uncovers gradients of temporal dynamics in the hippocampus and entorhinal cortex during rest and navigation. *BioRxiv*. https://doi.org/10.1101/2021.07.28.454036

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

Brunec, I. K., & Momennejad, I. (2022). Predictive Representations in Hippocampal and Prefrontal Hierarchies. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *42*(2), 299–312. https://doi.org/10.1523/JNEUROSCI.1327-21.2021

Brunec, I. K., Moscovitch, M., & Barense, M. D. (2018). Boundaries Shape Cognitive Representations of Spaces and Events. *Trends in Cognitive Sciences*, *22*(7), 637–650. https://doi.org/10.1016/J.TICS.2018.03.013

Butler, W. N., Hardcastle, K., & Giocomo, L. M. (2019). Remembered reward locations restructure entorhinal spatial maps. *Science*, *363*(6434), 1447–1452. https://doi.org/10.1126/science.aav5297

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

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage, 53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Diba, K., & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience, 10*(10), 1241–1242. https://doi.org/10.1038/nn1961

Dimsdale-Zucker, H. R., Ritchey, M., Ekstrom, A. D., Yonelinas, A. P., & Ranganath, C. (2018). CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-017-02752-1

Dupret, D., O'Neill, J., Pleydell-Bouverie, B., & Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature Neuroscience, 13*(8), 995–1002. https://doi.org/10.1038/nn.2599

Eichenbaum, H., & Cohen, N. J. (2014, August 20). Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? *Neuron*. https://doi.org/10.1016/j.neuron.2014.07.032

Ekstrom, A. D., & Ranganath, C. (2017). Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus*. https://doi.org/10.1002/hipo.22750

Eldar, E., Lièvre, G., Dayan, P., & Dolan, R. J. (2020). The roles of online and offline replay in planning. *ELife, 9*. https://doi.org/10.7554/eLife.56911

Ferbinteanu, J., & Shapiro, M. L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron, 40*(6), 1227–1239. https://doi.org/10.1016/S0896-6273(03)00752-9

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., … Dale, A. M. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex, 14*(1), 11–22. https://doi.org/10.1093/cercor/bhg087

Frank, L. M., Brown, E. N., & Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, *27*(1), 169–178. https://doi.org/10.1016/S0896-6273(00)00018-0

Garvert, M. M., Dolan, R. J., & Behrens, T. E. J. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *ELife*, *6*, 1–20. https://doi.org/10.7554/eLife.17086

Gauthier, J. L., & Tank, D. W. (2018). A Dedicated Population for Reward Coding in the Hippocampus. *Neuron, 99*(1), 179-193.e7. https://doi.org/10.1016/j.neuron.2018.06.008

George, D., Rikhye, R. V., Gothoskar, N., Guntupalli, J. S., Dedieu, A., & Lázaro-Gredilla, M. (2021). Clone-structured graph representations enable flexible learning and vicarious

evaluation of cognitive maps. *Nature Communications*. https://doi.org/10.1016/B0-12-369398-5/00341-8

Gershman, S. J. (2018). The successor representation: Its computational logic and neural substrates. *Journal of Neuroscience*, *38*(33), 7193–7200. https://doi.org/10.1523/JNEUROSCI.0151-18.2018

Gillespie, A. K., Astudillo Maya, D. A., Denovellis, E. L., Liu, D. F., Kastner, D. B., Coulter, M. E., … Frank, L. M. (2021). Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron*, *109*(19), 3149-3163.e6. https://doi.org/10.1016/j.neuron.2021.07.029

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1726–1731. https://doi.org/10.1073/pnas.0610561104

He, Q., Starnes, J., & Brown, T. I. (2022). Environmental overlap influences goal-oriented coding of spatial sequences differently along the long-axis of hippocampus. *Hippocampus*, (February 2021), 1–17. https://doi.org/10.1002/hipo.23416

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

Hsieh, L. T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. *Neuron*. https://doi.org/10.1016/j.neuron.2014.01.015

Huffman, D. J., & Stark, C. E. L. (2017). The influence of low-level stimulus features on the representation of contexts, items, and their mnemonic associations. *NeuroImage*, *155*(October 2016), 513–529. https://doi.org/10.1016/j.neuroimage.2017.04.019

Ito, H. T., Zhang, S. J., Witter, M. P., Moser, E. I., & Moser, M. B. (2015). A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. *Nature*, *522*(7554), 50–55. https://doi.org/10.1038/nature14396

Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(45), 12176–12189. https://doi.org/10.1523/JNEUROSCI.3761-07.2007

Kalm, K., Davis, M. H., & Norris, D. (2013). Individual sequence representations in the medial temporal lobe. *Journal of Cognitive Neuroscience*, *25*(7), 1111–1121. https://doi.org/10.1162/jocn_a_00378

Kaplan, R., Schuck, N. W., & Doeller, C. F. (2017). The Role of Mental Maps in Decision-Making. *Trends in Neurosciences*, *40*(5), 256–259. https://doi.org/10.1016/j.tins.2017.03.002

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Kubie, J. L., Levy, E. R. J., & Fenton, A. A. (2020). Is hippocampal remapping the physiological basis for context? *Hippocampus*, *30*(8), 851–864. https://doi.org/10.1002/hipo.23160

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M. B., & Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, *305*(5688), 1295–1298. https://doi.org/10.1126/science.1100265

Lu, Q., Hasson, U., & Norman, K. A. (2022). A neural network model of when to retrieve and encode episodic memories. *ELife*, *11*, e74445. https://doi.org/10.7554/eLife.74445

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markus, E., Qin, Y., Leonard, B., Skaggs, W., McNaughton, B., & Barnes, C. (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *The Journal of Neuroscience*, *15*(11), 7079–7094. https://doi.org/10.1523/jneurosci.15-11-07079.1995

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*(11), 1609–1617. https://doi.org/10.1038/s41593-018-0232-z

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McKenzie, S., Robinson, N. T. M., Herrera, L., Churchill, J. C., & Eichenbaum, H. (2013). Learning Causes Reorganization of Neuronal Firing Patterns to Represent Related Experiences within a Hippocampal Schema. *Journal of Neuroscience*, *33*(25), 10243–10256. https://doi.org/10.1523/jneurosci.0879-13.2013

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal Representation of Related and Opposing Memories Develop within Distinct,

Hierarchically Organized Neural Schemas. *Neuron*, *83*(1), 202–215. https://doi.org/10.1016/J.NEURON.2014.05.019

Mehta, M. R. (2001). Neuronal dynamics of predictive coding. *Neuroscientist*, *7*(6), 490–495. https://doi.org/10.1177/107385840100700605

Momennejad, I. (2020). Learning Structures: Predictive Representations, Replay, and Generalization. *Current Opinion in Behavioral Sciences*, *32*, 155–166. https://doi.org/10.1016/j.cobeha.2020.02.017

Muller, R. U., & Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *7*(7), 1951–1968. https://doi.org/10.1523/JNEUROSCI.07-07-01951.1987

Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage*, *103*, 130–138. https://doi.org/10.1016/j.neuroimage.2014.09.026

Mumford, J. A., Poline, J. B., & Poldrack, R. A. (2015). Orthogonalization of regressors in fMRI models. *PLoS ONE*, *10*(4), 1–11. https://doi.org/10.1371/journal.pone.0126255

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076

O'Keefe, J., Dostrovsky, J., & J. O'Keefe, J. D. (1971). Short Communications The hippocampus as a spatial map . Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–175. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5124915

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*.

O'Reilly, R. C., Ranganath, C., & Russin, J. L. (2021). The Structure of Systamaticity in the Brain. *ArXiv*, 1–22.

O'Reilly, R. C., Ranganath, C., & Russin, J. L. (2022). The Structure of Systematicity in the Brain. *Current Directions in Psychological Science*, *31*(2), 124–130. https://doi.org/10.1177/09637214211049233

Park, S. A., Miller, D. S., Nili, H., Ranganath, C., & Boorman, E. D. (2020). Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron*, *107*(6), 1226-1238.e8. https://doi.org/10.1016/j.neuron.2020.06.030

Patai, E. Z., Javadi, A., Ozubko, J. D., Callaghan, A. O., Ji, S., Robin, J., … Spiers, H. J. (2019). Hippocampal and Retrosplenial Goal Distance Coding After Long-term Consolidation of a Real-World Environment. *Cerebral Cortex*, 1–11. https://doi.org/10.1093/cercor/bhz015

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*(7447), 1–8. https://doi.org/10.1038/nature12112

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

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Sanders, H., Wilson, M. A., & Gershman, S. J. (2020). Hippocampal remapping as hidden state inference. *ELife*, *9*, 1–31. https://doi.org/10.7554/eLife.51140

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 773–786. https://doi.org/10.1098/rstb.2007.2087

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

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

Singmann, H., & Singmann, H., Bolker, B., Westfall, F., & A. F. (2016). afex: Analysis of Factorial Experiments.

Skaggs, W. E., & McNaughton, B. L. (1998). Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions. *Journal of Neuroscience*, *18*(20), 8455–8466. https://doi.org/10.1523/jneurosci.18-20-08455.1998

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

Summerfield, C., Luyckx, F., & Sheahan, H. (2020). Structure learning and the posterior parietal cortex. *Progress in Neurobiology*, *184*(September 2019). https://doi.org/10.1016/j.pneurobio.2019.101717

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A Map for Social Navigation in the Human Brain. *Neuron*, *87*(1), 231–243. https://doi.org/10.1016/j.neuron.2015.06.011

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

Turner, B. O., Mumford, J. A., Poldrack, R. A., & Ashby, F. G. (2012). Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage*, *62*(3), 1429–1438. https://doi.org/10.1016/j.neuroimage.2012.05.057

Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. J. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, *183*(5), 1249-1263.e23. https://doi.org/10.1016/j.cell.2020.10.024

Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods: Cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience*, *17*(8), 513–523. https://doi.org/10.1038/nrn.2016.56

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal Neurons Encode Information about Different Types of Memory Episodes Occurring in the Same Location. *Neuron*, *27*(3), 623–633. https://doi.org/10.1016/S0896-6273(00)00071-4

Yushkevich, P. A., Amaral, R. S. C., Augustinack, J. C., Bender, A. R., Bernstein, J. D., Boccardi, M., … Zeineh, M. M. (2015). Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards a harmonized segmentation protocol. *NeuroImage*, *111*, 526–541. https://doi.org/10.1016/j.neuroimage.2015.01.004

Acknowledgments: We thank Charles Lowell for his assistance with data collection. We thank Trevor Baer, Costin Tanase, Dennis Thompson, and the Imaging Research Center for their technical contributions. We thank Matthias Gruber, Zach Reagh, Halle Dimsdale-Zucker, Nichole Bouffard, Walter Reilly and the Dynamic Memory Lab for consultation on analysis and experimental design. We thank James Antony for feedback on the manuscript. This research was funded by a Royal Society and Wellcome Trust Sir Henry Dale Fellowship to AC (211200/Z/18/Z), a Multi-University Research Initiative grant N00014-17-1-2961 from the U.S. Office of Naval Research/Department of Defense awarded to C.R., and grant N00014-20-1-2578 from the U.S. Office of Naval Research/Department of Defense awarded to Randall O'Reilly and C.R. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the official views of the Office of Naval Research,

Author Contributions

Conceptualization, J.C.D., A.C., CR.; Methodology, J.C.D., A.C., D.H., S.A.P., CR.; Investigation, J.C.D., A.C.; Writing – Original Draft, J.C.D., A.C., C.R; Writing – Review & Editing, J.C.D., A.C., S.A.P., D.H., E.B., C.R.; Funding Acquisition, C.R.; Supervision, C.R.

Competing interests: The authors declare no competing financial interests.

Part 2:

To complete or separate: how the hippocampus prioritizes goal information during planning

WORKING TITLE

To complete or separate: how the hippocampus prioritizes goal information during planning

Authors

Jordan Crivelli-Decker, Yicong (Alan) Zheng, James Antony, Randall C. O'Reilly, Charan Ranganath

Introduction

Whether it is wayfinding in an unknown city or planning an optimal route through a crowded market, it is well-established that the hippocampus plays an important role in spatial navigation and memory (O'Keefe & Nadel 1978**,** Eichenbaum, 2017). Consistent with this role, numerous findings have shown that individual neurons within the hippocampus encode specific locations within a spatial context (O'Keefe and Dostrovsky, 1971). A critical behavior in everyday human life, planning, requires an organism to simulate the actions they need to perform to reach a given goal. It is possible that individual place memories could be chained together in order to construct a navigational plan, however, how the medial temporal lobe and hippocampus perform this function computationally is still unknown.

In artificial intelligence (AI) and decision-making research, planning is described as a model-based approach to control problems (Sutton & Barto 2018, Newell & Simon, 1972). These model-based methods use a policy for selecting actions on the basis of a prediction of an outcome given some internal model of the world (Dayan & Daw, 2008). One classic model for planning in complex decision spaces, heuristic search (Newell and Simon, 1972), states that a decision maker has a representation of the task at hand (e.g. an internal representation), then computes a plan by simulating and evaluating possible action sequences. Through a mnemonic lens, this process could be conceptualized by retrieving associations sequentially, using

knowledge of the transition probabilities between associations. Within this framework, it is clear that planning during navigation can be conceptualized as a model-based planning problem. Theories describing the processes that underlie navigation provide a systems level description of how the brain supports planning in this context (Schacter, Addis, & Buckner, 2007, Friston, 2010). However, the computational and neural principles of planning during navigation are poorly understood.

A large body of work from rodents and neuroimaging research has shown that the hippocampus is critical for retrieving sequential associations (Hsieh et al., 2014; Schapiro et al., 2016; Kalm, Davis, & Norris, 2013; Fortin, Agster and Eichenbaum 2002). Furthermore, evidence strongly suggests it is also involved in the prospective planning of navigational routes (Brown et al., 2014; Brown et al., 2016, Crivelli-Decker et al., 2021). Computational models have also demonstrated that the hippocampal circuit could retrieve non-overlapping memories from a partial cue (Kumaran & McClelland, 2012, Schapiro et al., 2017), via big loop recurrence within the hippocampal-entorhinal circuit. Whereby, the output of the hippocampal network is recirculated back into the network, enabling sequential retrieval of pairs of associations. Another class of models rely on fast timescale simulations of the future that are based on the agent's internal model of the world (Sutton, 1990; Mattar & Daw, 2018; Stoianov et al., 2018). These models query individual memories for states based on the agent's internal model, and are subsequently incorporated into its plan based on some policy (e.g to maximize reward). How these models query these states is an open area of debate, but empirical research suggests that individual place memories (indexed by place cells), are activated sequentially during planning to simulate possible futures (Johnson & Redish 2007; Pfeiffer and Foster 2013; Kay et al., 2020). This, along with evidence for recurrent activity in the hippocampal network in humans (Koster et al., 2018), suggest that the hippocampus may support planning by either sequentially retrieving overlapping associations, or via sequential activation of cell assemblies. Allowing an

agent to retrieve previously traversed paths and synthesize new paths using knowledge of the task (Wittkuhn et al., 2022).

These sequential chaining models provide a parsimonious account for how memory and planning can interact; however, they have several shortcomings. One potential issue with chaining models is how the system handles a so-called “broken link” in the chain, when the model incorrectly retrieves the next link in the chain. This is especially problematic when there are many highly overlapping associations with a single memory. How does the hippocampal system handle individual association that are highly overlapping and susceptible to interference?

There is a rich empirical and computational literature on how the hippocampus performs pattern separation (Leutgeb et al., 2007, Norman & O’Reilly, 2003). This key computation in the brain takes overlapping input patterns and transforms them into a very different output pattern, and is thought to play a key role in memory (Yassa & Stark, 2011; Cayco-Gajic & Silver, 2019). However, there has been very little work investigating how chaining models that utilize pattern separation might perform within a high interference scenario, like planning during navigation.

Another aspect of navigational memories that chaining models often neglect, is the impact of the goal itself. Recent evidence suggests that goals may exert a powerful influence on both spatial firing fields and representational patterns within the hippocampus. In rodents, hippocampal place cells are modulated during planning depending on the future goal of the animal (Ainge et al., 2007; Wood et al., 2000). Consistent with these findings, hippocampal activity patterns, measured with fMRI, carry information about future goal locations during planning and navigation (Brown et al., 2016; He et al., 2022; Crivelli-Decker et al., 2021. In addition, Crivelli-Decker et al. showed that while planning, hippocampal patterns had high similarity with trials that shared the same goal. This goal generalization effect seems to be at odds with many of the computational models that utilize pattern separation in the hippocampus

and it is unclear whether the hippocampus should pattern complete or pattern separate between plans that share many key features.

Rather than chaining individual associations together, it is also possible that input patterns into the hippocampal system are already structured to contain the most relevant parts of a navigational memory. This type of coding scheme would be highly advantageous when individual locations are incorporated in many different navigational plans. This process could be achieved by a network of regions that shape the inputs into the hippocampal system or directly modulate its internal dynamics. Some computational models have attempted to capture this process by demonstrating that highly overlapping memories could be initially separated by first retrieving the behaviorally relevant route features to guide future behavior (Zilli & Hasselmo, 2008a). This type of top-down control process would allow for retrieval of highly overlapping associations during high interference scenarios, in a way that prioritizes current contextual demands.

The present modeling work investigates possible mechanisms of goal-directed navigation within a CLS framework using two variations; a Chaining Model and a Schematic Model. The Chaining Model presented here attempts to simulate planning by retrieving pairs of overlapping associations in sequence. In contrast, the Schematic Model simulates planning by only retrieving the most critical aspects of a navigational memory. These models build upon prior CLS models (Ketz et al., 2013, McClelland, McNaughton & O'Reilly 2001) and includes neural network layers for hippocampal subfields DG, CA3, CA1. The model also includes a layer for the entorhinal cortex (EC), which is the primary input into the hippocampal formation. Importantly, this model utilizes a balance of error driven learning (EDL) and hebbian learning (Zheng et al., 2022) to retrieve and encode individual memories. This general framework has many advantages for capturing memory processes such as: how sub-regions with high inhibition, like DG, allow patterns to be separated and how an area with many recurrent connections (CA3), allows similar pattern to be completed given an incomplete input (Colgin,

Moser, & Moser, 2008; Marr, 1971; K. A. Norman & O'Reilly, 2003; Rolls & Kesner, 2006; Treves & Rolls, 1994).

Methods

Recurrent Neural Network of the Hippocampus

Learning in neural networks presented here occurs via the modification of weights between sending and receiving neurons, updated by two separate learning rules. One learning rule, Hebbian learning, states that changes in weights between two neurons are updated through repeated, simultaneous co-activation (Hebb, 1949). The other learning rule, error driven learning (EDL), is based on the idea that the network is constantly producing expectations based on the product of input activations and synaptic weights. Learning occurs by computing some cost function based on the difference between the model's prediction and the correct outcome (O'Reilly & Munakata, 2000). This model is implemented within the Leabra framework (Local, Error-driven, and Associative, Biologically Realistic Algorithm), which provides standard point-neuron rate-coded neurons, inhibitory interneuron-mediated competition and sparse, distributed representations, full bidirectional connectivity, and temporal-difference based error-driven learning dynamics (O'Reilly & Munakata, 2000; O'Reilly, Munakata, Frank, & Contributors, 2012). See https://github.com/emer/leabra for fully-documented equations, code, and several example simulations. The current models are based on the Theremin Model (Zheng et al., 2022), which utilizes a combination of both error-driven and hebbian learning, and further builds on previous complementary learning systems (CLS) models (Norman & O'Reilly, 2003), with additions for theta-phase dynamics (Ketz et al., 2013). The main difference in the present models are the modifications of the entorhinal cortex input layer into the network.

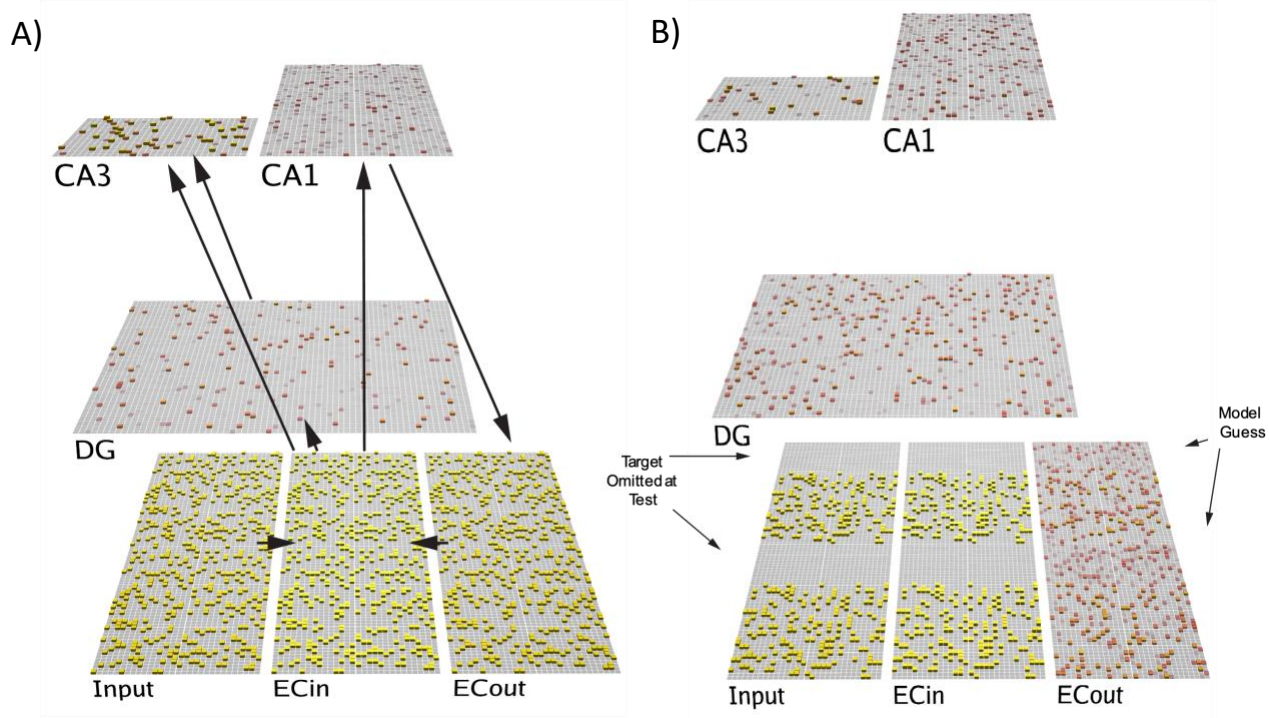

Figure 1: Primary Model Architecture and sample network training and testing procedure.

(A) Example training loop: Training patterns were input into the CLS model via ECin. Hidden layer connectivity is depicted with arrows. During training, the model learned to reconstruct the pattern presented to ECin in ECout.

(B) Example testing loop: During testing, the model was presented with a partially degraded pattern where "target" pools were omitted. Model performance was measured by how well the model reproduced "target" pool activity in ECout.

Figure 1 depicts the primary architecture of the hippocampal model presented here. Individual simulations differ as noted in the experimental paradigms section below. Our neural network is separated into the following layers:

1. An input layer comprising of goal pools, move pools, context pools, and state pools that each contained 100 neurons. These pools were organized into separate modules based to reflect how information from various cortical regions converges in the hippocampus (Eichenbaum et al., 2007**,** Witter, Doan, Jacobsen, Nilssen, & Ohara, 2017). This layer has one direct, one-to-one forward connection with ECin. Meaning, its inputs only drive activity in the ECin layer.

2. An ECin layer, which receives signals from the input layer and projects to DG and CA3 via broad, diffuse perforant path (PP) connections with a 25% chance of connection. This type of connection is important for driving conjunctive coding with DG and CA3. This random connectivity ensures that these subfields get a representative sample of the whole ECin layer.
3. A relatively large DG layer, which has a high level of inhibition, encouraging pattern separated representations. In both models presented here, DG has an inhibitory conductance multiplier of 3.2, which results in activity in about 1% of neurons. DG projects onto CA3 via mossy fiber projections (Henze, Wittner & Buzsaki, 2002), which have a strength multiplier of 4 and give it a stronger influence on CA3 activity than the direct projection for ECin discussed above.
4. A CA3 layer, which receives projections from both DG and ECin and projects onto itself (via recurrent collateral connection) and to CA1, with a strength multiplier of 2. Recurrent collaterals are theorized to be critical for pattern completion because an activated representation can retrieve its previously learned association within this layer (Marr, 1971; O'Reilly & McClelland, 1994)
5. A CA1 layer, which receives and compares inputs from ECin and CA3. This serves as a confluence for two hippocampal pathways and sends information back out of the hippocampus via ECout discussed below. The ECin -> DG -> CA3 -> CA1 -> ECOut pathway is commonly referred to as the trisynaptic pathway and is essential for pattern separation between highly similar inputs. The pathway from ECin -> CA1 -> ECout is commonly referred to as the monosynaptic pathway, which allows CA1 to directly encode and send activity from the hippocampus back into neocortex. These connections remain within pools, following their point-to-point anatomical connectivity patterns (Witter et al., 2017). This pathway can be thought of as an autoencoder function, which

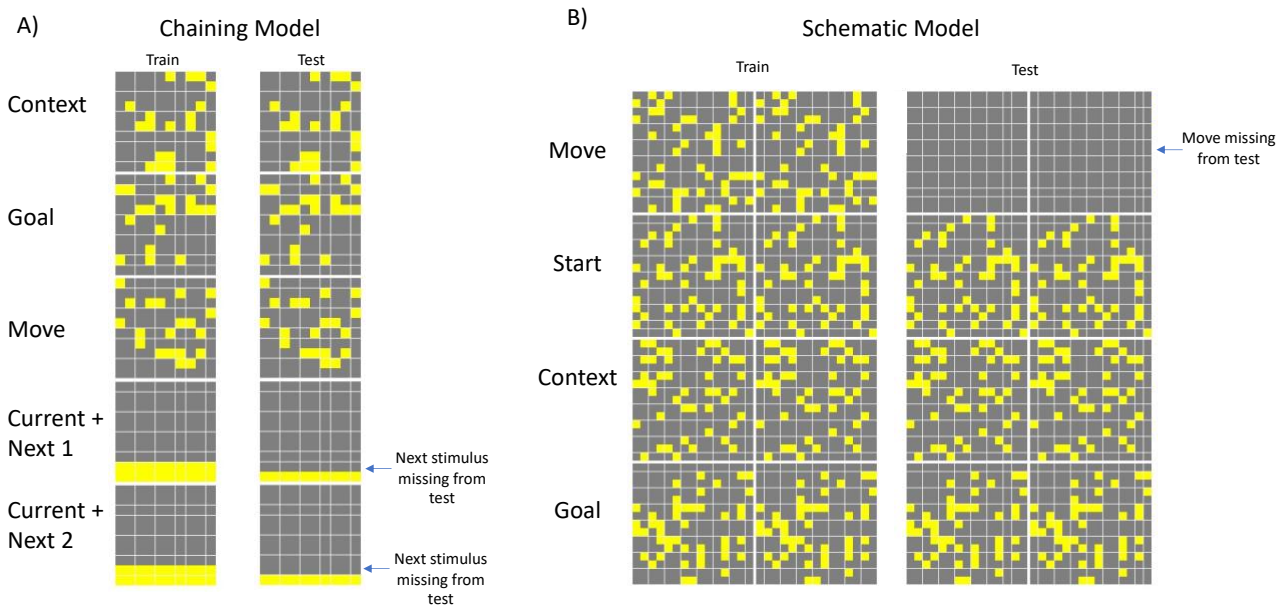

Figure 2 - Comparison of input patterns for the Chaining Model and Schematic Model

(A) Example input patterns during training and test for the Chaining Model: During training, the model was presented with five pools which consisted of the current context, current goal, current move, and two pools represented the pattern for the current and next stimulus in the sequence. During testing, the model was presented with the same pool structure as training, except for one key difference. The pools for current + next were degraded such that, they omitted the pattern information associated with the next item in the sequence. The model's performance was evaluated based on how well it could reproduce these omitted bits.

(B) Example input patterns during the training and testing for the Schematic Model: During training, the model was presented with ten pools. Pools consisted of the critical move for a given cue, start position, context, and goal location. During testing, the model was presented with the same pool structure as training except omitting the patterns representing the key move. The model's performance was evaluated based on how well it could reproduce these omitted bits.

translates pattern separated representations from the trisynaptic pathway back into the common reference frame for cortex.

6. An ECOut layer, which serves as the output of the hippocampus and serves as the networks guess during testing (Figure 1). It also serves as the input back into ECin which can result in different activations in successive cycles through the hippocampus (Kumaran & McClelland 2012; Schapiro et al., 2017). It is worth noting however that this was not a fully recurrent network because the number of cycles was fixed at 1.

Model training and testing followed four discrete phases resembling activity during the four quarters of the hippocampal theta rhythm (Ketz et al., 2013). The model learned via two

EDL mechanisms. In the first mechanism, the first three quarters constitute what are considered the minus phases, whereby the network produces an expected output based on its weights and input activations. The fourth and final quarter was the plus phase, whereby the target activation was provided from ECin → ECout and thereby learning occurred based on the difference between the network's prediction from the minus phases into ECout and the actual outcome. The first and fourth theta phases came during theta troughs, when CA1 was strongly influenced by ECin (Siegle & Wilson, 2014). Conversely, at theta peaks, CA1 was strongly influenced by CA3, which involved a guess based on activations and previously stored patterns. During the plus phase, ECin drove both CA1 and ECout activity, effectively clamping the correct answer in both EC layers and forcing weight adjustments in CA1. Therefore, across learning, ECout activity came to resemble ECin activity via the CA1 projection during the minus phases (without the direct ECin → ECout input). The second mechanism involved EDL in CA3 (Zheng et al., 2021). This error arose as a form of temporal difference learning between different pathways converging on CA3 neurons (Sutton & Barto, 2018): direct input from ECin (via the perforant path) and CA3 recurrent collateral activations arrived on CA3 neurons within the first quarter of the theta cycle, and critically, this input preceded signals from the multisynaptic ECin → DG → CA3 pathway (Yeckel & Berger, 1990). In our model, the minus phase constituted CA3 activity prior to DG inputs and the plus phase occurred when they arrived. Therefore, the pattern-separated DG activation acted as a teaching signal to correct the predicted pattern in CA3 based on perforant path + recurrent collateral activations alone (Kowadlo, Ahmed, & Rawlinson, 2019).

Experimental paradigms used in simulations

Our simulation efforts were focused on a state-space navigation paradigm that required participants to learn about two distinct contexts that consisted of animals that were linked in a deterministic sequence structure (Crivelli-Decker et al., 2021). Each context consisted of the

same animals organized in a plus maze topology, but the relationship between animals across contexts were mirror-reversed and then shifted counter-clockwise (Figure 3A). To model this task, we used individual binary patterns that were used to symbolize each of the animals in both contexts (Figure 2, Figure 3A). Note that the same state identities were used across contexts but were differentiable by using several other input pools into the network. In both models, we also included binary pools for Move, Context, and Goal. During training, we presented the model with collections of binary patterns (pools) that the model used to reconstruct (Figure 2). During testing, the model was presented with a degraded pattern where either some of the pools were empty or certain bits in the pattern were corrupted. Model performance was calculated by counting the number of completion bits (not presented at input). If the model was able to construct 50% of the completion bits, it was counted as a successful retrieval.

Results

We simulated an experiment (Crivelli-Decker et al., 2021) where subjects were required to retrieve memories that had a high amount of overlap both within and across contexts. Briefly, in this experiment, subjects learned about how to navigate to 4 goal animals in two contexts (Figure 3A). Critically, the items in each context were visually identical but were organized in a different orientation across contexts. A key finding from this work was that, during planning, within context routes that converge upon a similar goal show higher pattern similarity within the hippocampus, when compared to routes that do not share a goal. This experiment is challenging for a model that heavily relies on pattern separation because states (individual items) are shared across sequences (states organized in specific ordered positions) and contexts, and thus have high levels of interference. One state that is especially susceptible, is the middle position of the maze, because it is shared with every sequence (position 3).

An open question in both the model-based planning and memory literature is, whether the hippocampus actually encodes a map-like representation or a model of the world (Behrens et al., 2018; Ekstrom and Rangnath, 2018). If map-like representations arise in the hippocampus de-novo, one might expect that learning about an environment via pairwise associations would allow the hippocampus to learn about the transition structure of the environment. As a result, during planning, the hippocampus could utilize this learned structure to effectively chain these associations together. However, results from Crivelli-Decker et al. suggest that, during planning the hippocampus retrieves the specific information needed to reach a given goal. This points to the idea that, instead of the hippocampus learning solely about the transition structure, this structure information comes from interconnected cortical regions (Ritchey & Ranganath, 2012; Summerfield et al., 2020). We began by constructing a model that plans by chaining individual associations between states in two distinct contexts.

Simulation 1 – Representing individual associations between states fails in high interference situations: The Chaining Model

To simulate planning in this experiment, we trained a model with sequences of binary patterns that captured the sequences that participants experienced in Crivelli-Decker et al. Each animal in both mazes was assigned an arbitrary state number which received a unique binary pattern that would be used during the simulations (Figure 3A). Each binary pattern presented to the model represented an individual position in a sequence. A pattern consisted of a structured set of pools that contained the current context of the sequence, the current goal for the sequence, the move a participant was required to make at that position, and the current and next state in the sequence. These patterns were presented in sequence such that during training, the model would learn about the pool structure of position 1 then, position 2, continuing to position 4 (Figure 3B). A sequence consisted of a set of states organized in specific positions. The

sequences used here are identical to those used in Crivelli-Decker et al., At each position, the model learned to reproduce the input patterns presented to ECin in Ecout. After a given sequence was complete, we presented the next sequence, in order, until all positions in all sequences were repeated once for a given epoch (16 total epochs). In total, the model learned 96 associations, 48 for each context, 12 sequences, with 4 positions each. During testing, we presented the model with a degraded pattern where the pool for the next state in the sequence was missing. The model's performance was evaluated based on how well the model could reproduce these missing bits.

To evaluate whether this type of chaining could support planning in this task, we examined the model's ability to retrieve the next item in the sequence on epoch-by-epoch basis. Figure 3C shows that the chaining model was able to achieve ceiling level performance at position 1 and 4 after the training loop. This is notable because these positions have the smallest number of shared associations. In contrast, the model struggled to reproduce the next item in the sequence in both positions 2 and positions 3, displaying especially poor performance at position 2, which continued to get worse during training. This suggests that, at least in situations of low interference, models that rely on pattern separation may be able capture planning behavior.

To better understand why the model was unable to reach peak performance at positions 2 and 3, we analyzed the types of errors that the model was making. During test, we examined the activation of the completion pools (not presented at input), to determine what states the model was incorrectly predicting. If the model produced 50% of the units in a non-completion pool, this was coded as the model's guess for the state. The model tended to make most of its errors when it tried to produce state 2, 3, and 4 (Figure 3D, Left panel), which corresponded to positions 2 and 3 in the sequence (Figure 3C). This is notable because these states are associated with a large number of different sequences. Examining positions 2 and 3 more closely, we found that at position 2, the model erroneously reproduces the item that immediately

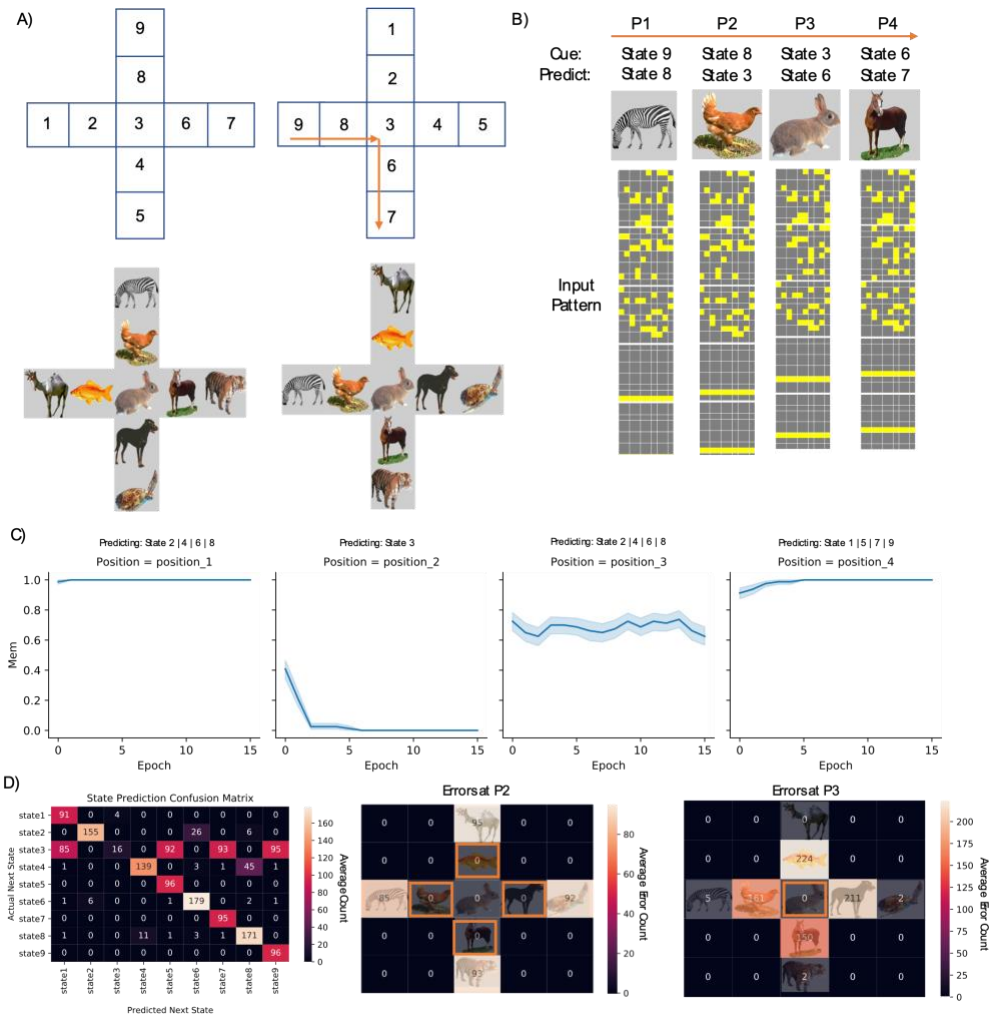

Figure 3 – Simulation approach and results from the Chaining Model

(A) Arbitrary state assignment to each of the animals in the mazes used in Crivelli-Decker et al. *Top panel*: State assignment diagram. Arrow in orange depicts an example sequence of inputs presented to the model. *Bottom Panel:* Mazes presented to subjects in Crivelli-Decker et al.

(B) Example sequence of input patterns into the hippocampal model. At each position the model was presented with a binary pattern cue and was required to predict the next item in the sequence.

(C) Average behavioral performance for the chaining model separated by position. The model was able to accurately reproduce the next state in the sequence with near perfect accuracy at positions 1 and 4. Positions 2 and 3 the model struggled to reproduce the correct result.

(D) Error analysis for the chaining model. *Left panel:* Confusion matrix illustrating the types of errors made by the model. The diagonal of this matrix represents conditions where the model was correct in its predictions. Errors are clustered around when the model is trying to predict state 2,3, and 4. *Middle panel:* Errors at position 2. Orange boxes represent all possible current locations of the model at position 2. Overlaid on top of the maze is an average count of the predictions of the model when presented with the stimulus in the orange boxes. This shows the models inability to correctly predict state 3, the highest interference state. *Right panel:* Errors at position 3. The orange box represents the current state presented to the model. Errors are uniformly distributed on each of the arms, indicating that when presented with state 3 the model incorrectly tries to predict an out of sequence arm.

preceded it in the sequence. For example, in a sequence state 1 -> state 2-> state 3 -> state->

4, if presented with state 2, it would predict state 1. In addition, at position 3, the model performed poorly after training, preferring to predict an out-of-sequence state (Figure 3D, Right panel). Taken together, these results demonstrate that in situations where individual associations have high levels of interference, simply chaining associations is not enough to successfully pattern complete using a standard CLS model.

Simulation 2 – Structured inputs into hippocampus drive goal generalization via shared pattern information: The Schematic Model

In the previous section we demonstrated that a simple version of a chaining model using the CLS framework was unable to effectively perform the maze task. The results from Crivelli-Decker et al. suggest that, during planning and navigation, behaviorally relevant points in a sequence are prioritized within the hippocampus. One interpretation of this result is that, if an agent understands the overall task structure (e.g. an internal model or collection of rules), then during planning all they need to recall is the action to be taken at the center point in the maze. It is possible that the hippocampus receives a structured input pattern that already contains information about the structure of the task, simplifying pattern completion operation for the hippocampus (Ritchey & Ranganath 2012; Wittier et al., 2017; Summerfield et al., 2020). To capture this effect, we reasoned that during the planning phase, inputs into the hippocampus would contain a more schematic representation of a memory, only representing the critical features of the memory that distinguish it from other similar navigational memories.

To simulate this hypothesis, we exposed the model to pools that contained the starting state, goal state, context, and the critical move to reach the goal for that sequence in that context (Figure 2 & Figure 4B). A key difference between this model and the chaining model, is that each sequence only has one set of input patterns associated with it. During testing, the model was presented with a partial cue where the critical move was omitted and the model was

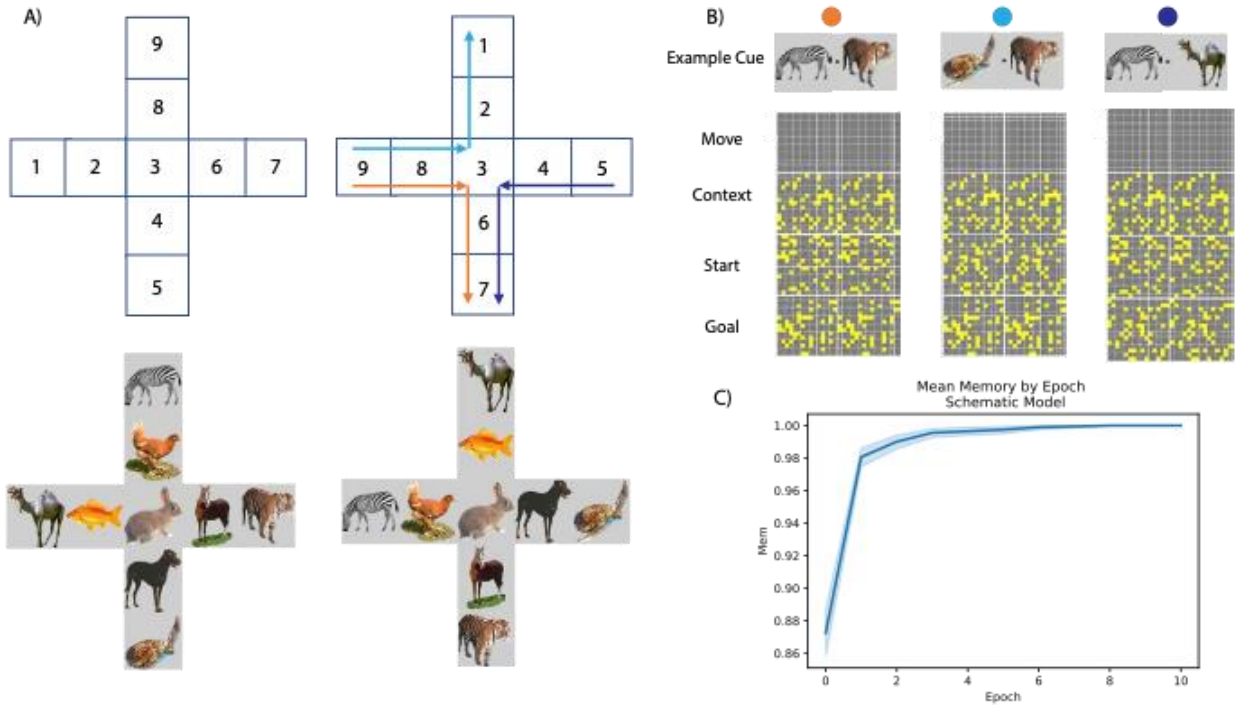

Figure 4 – Simulation Approach and behavioral results from Schematic Model

(A) T*op Panel:* Diagrams depicting the state to animal mapping used in the experiment. Colored arrows indicate different types of cues that either converge or diverge to different goals.

(B) Examples of input patterns during test. Three cues are presented that either converge or diverge on the same goal. Note that the binary patterns for the goal pools are the same in the first and second columns.

(C) Behavioral performance of the model separated by epoch.

required to predict the decision for the cue. We hypothesized, using input patterns structured like this would reduce interference and help the model pattern complete, improving model performance. We performed this simulation 100 times, each time producing a set of randomly generated binary input patterns for each session, in order to better simulate individual differences across subjects (Schapiro et al., 2017). Model performance was then analyzed on a per-run basis, where each run is analogous to a subject.

In contrast to the Chaining model, the Schematic model learned more quickly and was able to reach peak performance after (Average epochs until asymptotic = 5.76, SD = 0.68). In order to better understand the types of representations that this model was learning, we examined the representational similarity between individual trials within the hidden layers of the network after the training loop was complete.

This revealed a pattern of results that resembled the within context goal similarity effects observed in Crivelli-Decker et al (Figure 5). Within the same context, all subfields showed higher pattern similarity for converging than diverging sequences, though this effect was strongest for CA1 and DG (CA3: $T_{99} = 141.57$, $p < 0.001$; CA1: $T_{99} = 173.17$, $p < 0.001$; DG: $T_{99} = 210.84$, $p < 0.001$). Interestingly, all three subfields also displayed a significant effect of context (*Main effect of context*- CA3: $F(1,99) = 29068.39$, $p < 0.001$; CA1: $F(1, 99) = 153284.68$, $p < 0.001$; DG: $F(1, 99) = 12261.9361$). These results suggest that, given a structured input pattern, our

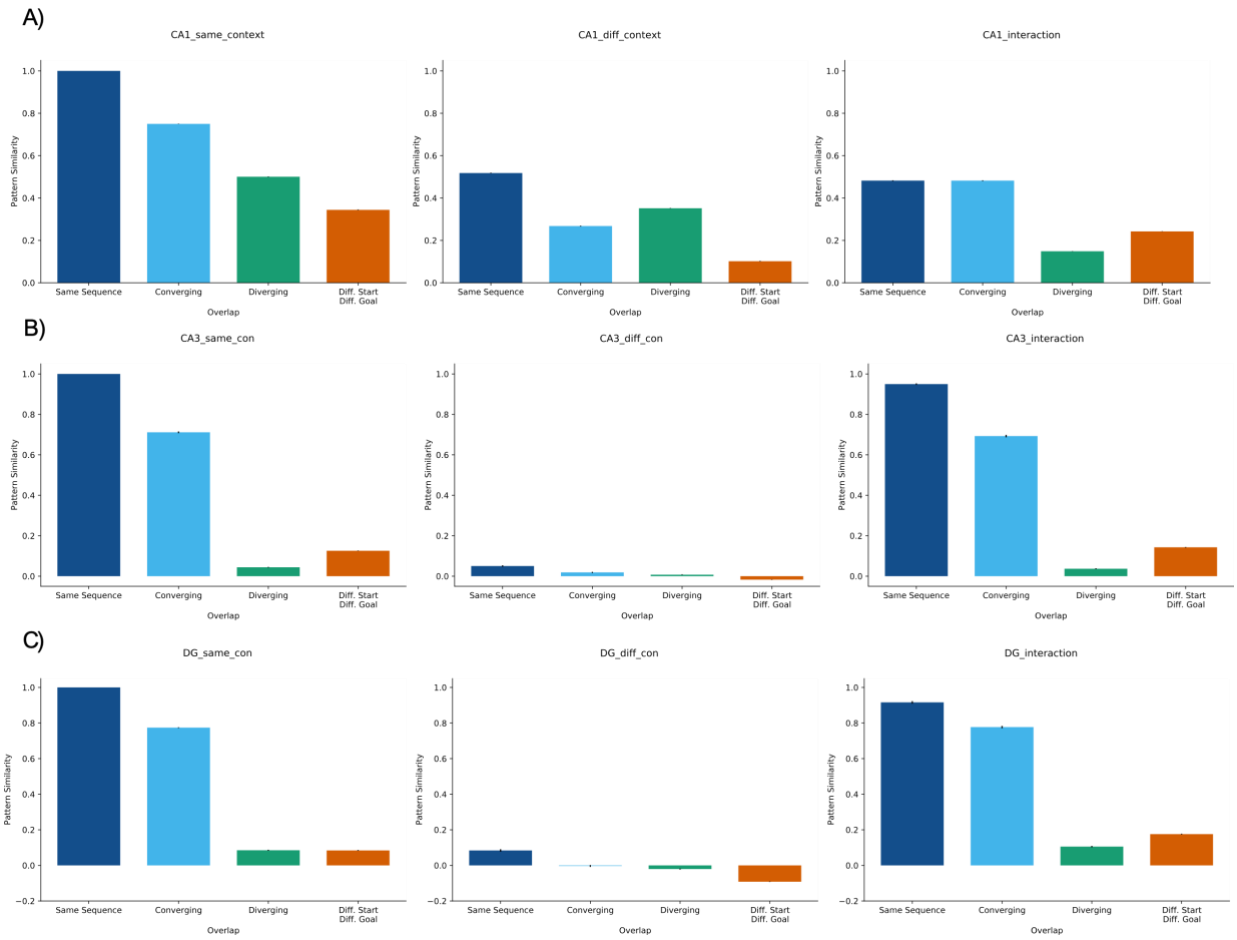

Figure 5 – Pattern similarity results from Schematic Model

(A) Pattern similarity results for the CA1 showing different overlap conditions separated by context. Interaction plots depict the between context overlap effect. Values greater than 0 indicate that pattern similarity is higher in the same context, whereas values less than 0 indicates that pattern similarity is higher in a different context.
(B) Same as A) but depicting CA3
(C) Same as A) but depicting DG

network is successfully able to pattern complete even in situations of high interference. Taken together, this highlights the need for extra-hippocampal coordination, either at the input level or through modulation of hippocampal dynamics, to retrieve sequential plans.

Discussion

Within a biologically plausible model of the HC-EC network, we simulated two possible computational mechanisms that may underlie planning. First, we ran simulations a Chaining Model, using incomplete pairs of associations as inputs into the network, where it was required to pattern complete the next state in the sequence. Using a task from Crivelli-Decker et al. we found that this model was highly susceptible to interference at positions (P2 and P3). These positions required the model to pattern separate between a large number of possible next steps, highlighting some major drawbacks of a simple chaining network when applied to navigation. Next, we adjusted the input pattern structure such that information entering the hippocampal network was already structured to only contain the most relevant information for retrieving the correct memory (Schematic Model). Using this pool structure, the model was able to successfully perform the navigation task from Crivelli-Decker et al. When analyzing pattern similarity within the subfields of the hippocampal network, we found that CA3 and DG displayed a goal generalization effect akin to their findings. Taken together, this highlights the importance of structured inputs into the hippocampal network and suggest that, without some top-down control, the hippocampus may struggle in situations with high interference.

Models that attempt to account for associative processes in memory have a long history in psychology and neuroscience, and first were discussed in the early experimental work of Herman Ebbinghaus (Ebbinghaus, 1913). The central idea of this framework is that when two items become active simultaneously, the items' representations become associated such that activation of one will evoke the other. Within the context of navigation and memory, there is

active debate regarding whether chains of associations support transitive inference or if they are supported by recurrent activity within the hippocampal network (Kumaran & McClelland, 2012, Schapiro et al., 2017). One theory of recurrence, the “big loop” theory, states that the hippocampus can form chains from non-overlapping associations via recirculation of hippocampal output as a new input (Kumaran & McClelland 2012). Empirical work suggests that this may be supported in part by anatomical and functional connections between deep and superficial layers of entorhinal cortex which correspond to output and input to the hippocampus, respectively (Koster et al., 2018, Witter et al., 2017). For example, Koster et al., found that in a transitive inference there was evidence of reactivation of the unseen bridging element (e.g. B in an AC pair) in both superficial and deep layers of EC**.** Superficial and deep layers of entorhinal cortex described in this work roughly correspond to ECin and ECout in our network. It is possible that via this recursion mechanism, the hippocampus could form long chains of associations that support planning. However, our results suggest that in situations of heightened interference, this type of recurrence would likely suffer when the system is unable to pattern complete (Figure 3).

There are several notable differences between our chaining model and the model presented in Kumaran et al. One key difference is that, our model was not fully recurrent and did not propagate outputs back into the network. This is a simplification of the recursive chaining mechanism and likely paints our chaining model in a more favorable light. If our model was allowed to recur outputs back into the network, it would likely have performed even worse at points of high interference, highlighting this mechanism’s susceptibility to a “broken link” in the chain. Another model that utilized recursion, Schapiro et al., 2017, found that when using a similar architecture to the one presented here, the hippocampus was able to learn about temporal communities through statistical learning. However, their input patterns were structured in a way where each pool was only activated by a single item. This type of input pattern could be thought of as a hybrid between the Chaining model and Schematic model presented here. It

is possible that if using an input pattern structure similar to the one used in Schapiro et al., our chaining model may have performed better at the task. Future work should further investigate if a fully recurrent chaining network is an effective model-based planner.

Computational attempts to capture goal-directed navigation and planning, have pointed to the need for a separate representation of the future that is used by the system to navigate (Burgess, Recce, & O'Keefe, 1994, Foster, Morris, & Dayan 2000). The prefrontal cortex has well established connectivity with entorhinal cortex (Witter et al., 2017) and has also been shown to have an indirect connection with CA1 via the nucleus reunions of the thalamus (Ito et al., 2015). It is likely that other brain areas that are interconnected with the hippocampal circuit participate in the control over hippocampal inputs and contribute to navigational planning. For example, representing goal information (Hok et al., 2005; Ito et al., 2015; Javadi et al., 2017), in navigation tasks. In the context of our work, prefrontal cortex could provide pattern information to ECin in the form of the goal pool presented to our models.

It is possible that via top-down control over hippocampal function, interference between individual associations could be reduced. For example, a system that is able to strategically control memory retrieval or encoding in a way that minimizes interference would huge benefits operating in high interference scenarios. One example of a computational model that relies on strategic deployment of past experiences comes from Lu, Hasson, and Norman (2022). Their simulations showed that it was computationally advantageous for their network if it prioritized encoding and retrieval to moments of high uncertainty. This type of control may be useful for chaining models to minimize interference between chains. Within the context of the Schematic Model, it would also help prioritize encoding of information related to position 3, a critical point of uncertainty.

In the Schematic Model, we used inputs that were already structured in a way that was maximally useful for the network. The central assumption of this model is that knowledge of the rules, transition structure, and goal information comes from interconnected cortical regions

(Ritchey & Ranganath, 2012; Summerfield et al., 2020). In this scenario, our model only needed to pattern complete the most arbitrary aspect of the task, which is the correct action at the decision point (position 3). This is in contrast to the Chaining Model, which also needed to encode aspects of the transition structure, via pairwise associations between items. This type of structure can be thought of as establishing a set of inductive biases that the model can use to perform the task. In many situations, an agent with an appropriate understanding of the structure of the task could benefit by preferentially encoding and retrieving key moments that minimize interference between memories (O'Reilly, Russin & Ranganath, 2022, Zilli & Hasselmo, 2008a). By incorporating this set of inductive biases into our network, we were able to replicate the main pattern similarity findings in Crivelli-Decker et al, showing that hippocampal activity patterns differentiated similar routes based on a goal. It is possible either via inductive biases or top-down modulation from other brain regions contributed to their unique set of results. Future work should utilize additional modules that represent top-down control over both hippocampal inputs and internal hippocampal dynamics to better capture the anatomy and empirical findings in the literature.

References:

Ainge, J. A., Tamosiunaite, M., Woergoetter, F., & Dudchenko, P. A. (2007). Hippocampal CA1 place cells encode intended destination on a maze with multiple choice points. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(36), 9769–9779. https://doi.org/10.1523/JNEUROSCI.2011-07.2007

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

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

Brown, T. I., Hasselmo, M. E., & Stern, C. E. (2014). A High-resolution study of hippocampal and medial temporal lobe correlates of spatial context and prospective overlapping route memory. *Hippocampus*, *24*(7), 819–839. https://doi.org/10.1002/hipo.22273

Burgess, N., Recce, M., & O'Keefe, J. (1994). A model of hippocampal function. *Neural Networks*. Netherlands: Elsevier Science. https://doi.org/10.1016/S0893-6080(05)80159-5

Cayco-Gajic, N. A., & Silver, R. A. (2019). Re-evaluating Circuit Mechanisms Underlying Pattern Separation. *Neuron*, *101*(4), 584–602. https://doi.org/10.1016/j.neuron.2019.01.044

Colgin, L. L., Moser, E. I., & Moser, M. B. (2008). Understanding memory through hippocampal remapping. *Trends in Neurosciences*, *31*(9), 469–477. https://doi.org/10.1016/j.tins.2008.06.008

Crivelli-Decker, J., Clarke, A., Park, S. A., Huffman, D. J., Boorman, E. D., & Ranganath, C. (2021). Goal-centered representations in the human hippocampus, bioRxiv 1–45. https://doi.org/10.1101/2021.08.18.456881

Eichenbaum, H. (2017, August 30). On the Integration of Space, Time, and Memory. *Neuron*. Elsevier. https://doi.org/10.1016/j.neuron.2017.06.036

Ekstrom, A. D., & Ranganath, C. (2018). Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus*, *28*(9), 680–687. https://doi.org/10.1002/hipo.22750

Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, *5*(5), 458–462. https://doi.org/10.1038/nn834

Foster, D., & Wilson, M. A. (2007). Hippocampal Theta Sequences. *Hippocampus*, *17*, 1093–1099. https://doi.org/10.1002/hipo

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

He, Q., Starnes, J., & Brown, T. I. (2022). Environmental overlap influences goal-oriented coding of spatial sequences differently along the long-axis of hippocampus. *Hippocampus*, (February 2021), 1–17. https://doi.org/10.1002/hipo.23416

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory. The organization of behavior; a neuropsychological theory.* Oxford, England: Wiley.

Henze, D. A., Wittner, L., & Buzsáki, G. (2002). Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. *Nature Neuroscience*, *5*(8), 790–795. https://doi.org/10.1038/nn887

Hok, V., Save, E., Lenck-Santini, P. P., & Poucet, B. (2005). Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(12), 4602–4607. https://doi.org/10.1073/pnas.0407332102

Hsieh, L. T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. *Neuron*. https://doi.org/10.1016/j.neuron.2014.01.015

Ito, H. T., Zhang, S. J., Witter, M. P., Moser, E. I., & Moser, M. B. (2015). A prefrontal-thalamo-hippocampal circuit for goal-directed spatial navigation. *Nature*, *522*(7554), 50–55. https://doi.org/10.1038/nature14396

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

Johnson, A., & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(45), 12176–12189. https://doi.org/10.1523/JNEUROSCI.3761-07.2007

Kalm, K., Davis, M. H., & Norris, D. (2013). Individual Sequence Representations in theMedial Temporal Lobe. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn

Kay, K., Chung, J. E., Sosa, M., Schor, J. S., Karlsson, M. P., Larkin, M. C., … Frank, L. M. (2020). Constant Sub-second Cycling between Representations of Possible Futures in the Hippocampus. *Cell*, *180*(3), 552-567.e25. https://doi.org/10.1016/j.cell.2020.01.014

Ketz, N., Morkonda, S. G., & O'Reilly, R. C. (2013). Theta Coordinated Error-Driven Learning in the Hippocampus. *PLoS Computational Biology*, *9*(6). https://doi.org/10.1371/journal.pcbi.1003067

Koster, R., Chadwick, M. J., Chen, Y., Berron, D., Banino, A., Düzel, E., … Kumaran, D. (2018). Big-Loop Recurrence within the Hippocampal System Supports Integration of Information across Episodes. *Neuron*, *99*(6), 1342-1354.e6. https://doi.org/10.1016/j.neuron.2018.08.009

Kowadlo, G., Ahmed, A., & Rawlinson, D. (2019). AHA! an "Artificial Hippocampal Algorithm" for Episodic Machine Learning. Retrieved from http://arxiv.org/abs/1909.10340

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, *119*(3), 573–616. https://doi.org/10.1037/a0028681

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

Lu, Q., Hasson, U., & Norman, K. A. (2021). When to retrieve and encode episodic memories: a neural network model of hippocampal-cortical interaction. *BioRxiv*, 2020.12.15.422882. Retrieved from https://www.biorxiv.org/content/10.1101/2020.12.15.422882v2%0Ahttps://www.biorxiv.org/content/10.1101/2020.12.15.422882v2.abstract

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *262*(841), 23–81. https://doi.org/10.1098/rstb.1971.0078

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*(11), 1609–1617. https://doi.org/10.1038/s41593-018-0232-z

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review.* [Washington, etc.]: American Psychological Association.

Newell, A., & Simon, H. A. (1972). Human problem solving. *Human Problem Solving.* Oxford, England: Prentice-Hall.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*. Norman, Kenneth A.: Princeton U, Dept of Psychology, Green Hall, Princeton, NJ, US, 08544, knorman@princeton.edu: American Psychological Association. https://doi.org/10.1037/0033-295X.110.4.611

O'Keefe, J., Dostrovsky, J., & J. O'Keefe, J. D. (1971). Short Communications The hippocampus as a spatial map . Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–175. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5124915

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*.

O'Reilly, R. C., & Munakata, Y. (2000). Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain.* Cambridge, MA, US: The MIT Press.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, *4*(6), 661–682. https://doi.org/10.1002/hipo.450040605

O'Reilly, R. C., Ranganath, C., & Russin, J. L. (2021). The Structure of Systamaticity in the Brain. *ArXiv*, 1–22.

Pezzulo, G., Kemere, C., & van der Meer, M. A. A. (2017). Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition. *Annals of the New York Academy of Sciences*, *1396*(1), 144–165. https://doi.org/10.1111/nyas.13329

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*(7447), 1–8. https://doi.org/10.1038/nature12112

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, *13*(10), 713–726. https://doi.org/10.1038/nrn3338

Richard S., S., & Barto, A. G. (2018). *Reinforcement Learing - An Introduction*.

Rolls, E. T., & Kesner, R. P. (2006). A computational theory of hippocampal function, and empirical tests of the theory. *Progress in Neurobiology*, *79*(1), 1–48. https://doi.org/10.1016/j.pneurobio.2006.04.005

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657–661. https://doi.org/10.1038/nrn2213

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

Siegle, J. H., & Wilson, M. A. (2014). Enhancement of encoding and retrieval functions through theta phase-specific manipulation of hippocampus. *ELife*, *2014*(3). Retrieved from https://elifesciences.org/content/3/e03061

Stoianov, I. P., Pennartz, C. M. A., Lansink, C. S., & Pezzulo, G. (2018). Model-based spatial navigation in the hippocampus-ventral striatum circuit: A computational analysis. *PLoS Computational Biology*, *14*(9), 1–28. https://doi.org/10.1371/journal.pcbi.1006316

Summerfield, C., Luyckx, F., & Sheahan, H. (2020). Structure learning and the posterior parietal cortex. *Progress in Neurobiology*, *184*(September 2019). https://doi.org/10.1016/j.pneurobio.2019.101717

Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGART Bulletin*, *2*(4), 160–163. https://doi.org/10.1145/122344.122377

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–391. https://doi.org/10.1002/hipo.450040319

Witter, M. P., Doan, T. P., Jacobsen, B., Nilssen, E. S., & Ohara, S. (2017). Architecture of the entorhinal cortex a review of entorhinal anatomy in rodents with some comparative notes. *Frontiers in Systems Neuroscience*, *11*(June), 1–12. https://doi.org/10.3389/fnsys.2017.00046

Wittkuhn, L., Chien, S., Hall-McMaster, S., & Schuck, N. W. (2021). Replay in minds and machines. *Neuroscience and Biobehavioral Reviews*, *129*, 367–388. https://doi.org/10.1016/j.neubiorev.2021.08.002

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal Neurons Encode Information about Different Types of Memory Episodes Occurring in the Same Location. *Neuron*, *27*(3), 623–633. https://doi.org/10.1016/S0896-6273(00)00071-4

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yeckel, M. F., & Berger, T. W. (1990). Feedforward excitation of the hippocampus by afferents from the entorhinal cortex: Redefinition of the role of the trisynaptic pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(15), 5832–5836. https://doi.org/10.1073/pnas.87.15.5832

Zheng, Y., Liu, X. L., Nishiyama, S., Ranganath, C., & Randall, C. (2021). Correcting the Hebbian Mistake : Toward a Fully Error-Driven Hippocampus, *2067*(Mdm), 1–22.

Zilli, E. A., & Hasselmo, M. E. (2008). Analyses of Markov decision process structure regarding the possible strategic use of interacting memory systems. *Frontiers in Computational Neuroscience*, *2*(DEC), 1–13. https://doi.org/10.3389/neuro.10.006.2008