Planning in the Hippocampus: Linking Actions and Outcomes to Guide Behavior

Sarah Jo C. Venditto
(venditto@princeton.edu)
Princeton Neuroscience Institute,
Princeton University

Kevin J. Miller
(kevinjmiller@deepmind.com)
Google DeepMind & University College London

Nathaniel D. Daw
(ndaw@princeton.edu)
Princeton Neuroscience Institute,
Princeton University

Carlos D. Brody
(brody@princeton.edu)
Princeton Neuroscience Institute & HHMI,
Princeton University

Abstract

Planning requires an internal model of the world that can be flexibly utilized to link actions and subsequent consequences across time and space. The hippocampus, often referred to as a "cognitive map," is known for encoding the location of an animal within complex environments by representing salient states, both spatial and non-spatial. These representations can extend to non-local states, making them well-suited to support this internal action-outcome model. While the hippocampus has been causally linked to planning in both humans and rodents, how hippocampal representations carry out this function is poorly understood. To address this, we record from dorsal hippocampus while rats perform a multi-step reward-guided task that employs probabilistic transitions between actions and outcomes, the rat two-step task, which has been shown to reliably elicit planning. We find that hippocampal activity encodes the task space and exhibits "splitter cells" that differentiate similar positions based on preceding choice, providing distinct representations for each combination of choice and outcome. In-between trials, we find oscillating representations that encode the visited outcome paired with both possible choices; however, overall choice encoding is biased towards upcoming choices with a model-based dependence on reward and probabilistic transition.

Keywords: planning; hippocampus; model-based reinforcement learning; model-based credit assignment

Hippocampal activity tiles task space

Rat two-step task and electrophysiological recordings. Rats performed a multi-step probabilistic decision task for liquid reward (Figure 1), known as the two-step task. Rats update choices during a first step (ii) in light of outcomes (v) and rewards (vi) observed at a second step. One outcome will have a higher chance of reward than the other, and the better outcome will reverse unpredictably throughout the session (vii). Choices are linked probabilistically to outcomes (iii), where each has a high probability of leading to one of the outcomes (common) and a low probability of leading to the opposite outcome (rare). Previous work demonstrated that rats solve this task using primarily a model-based (MB) strategy, where animals' choices were influenced by the interaction between reward and transition (Miller, Botvinick, & Brody, 2017). After rats were fully trained, each were implanted bilaterally with neuropixels 2.0 probes targeting layer CA1 of dorsal hippocampus, where we recorded from 15,361 task-active units from 131 sessions across 4 rats.

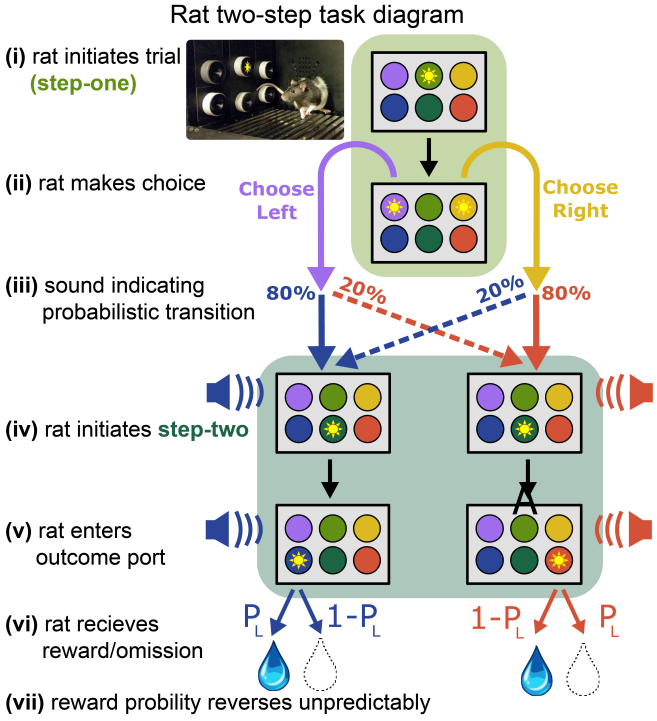

Figure 1: Rat two-step task description.

Hippocampal activity maps distinguish choice history and trial type. In order for the hippocampus to support an internal-action outcome model, it needs to represent states of the task. We measure average firing rates separately across each of four trial types defined by each combination of choice port (left, L1, and right, R1) with outcome port (left, L2, and right, R2). Each trial is represented as a linear trajectory between all ports visited during the trial. A summary of all task-selective units can be seen in Figure 2A, where units are rank sorted by preferred port (C1,L1,R1,C2,L2,R2) and trial type (L1-L2,L1-R2,R1-L2,R1-R2), and colored according to preferred port. We find selectivity across all positions, where many units differentiate the same position dependent on choice history. For example, Figure 2Bi shows a unit fir-

ing at second-step initiation (C2) only after left choice. Figure 2Bii shows an example unit firing during reward after a right-outcome port (R2) only following a right choice, i.e. only during a common transition. Figure 2Biii shows an example unit firing during reward following a left-outcome (L2), but only after a right choice was made, i.e. only during a rare transition. All of these units are examples of what is known as splitter cells, a common phenomena in hippocampal encoding, where units only fire in a particular position in its preferred context (Duvelle, Grieves, & van der Meer, 2023). The existance of second-step splitter cells suggests that trial types are differentiated by separable activity maps.

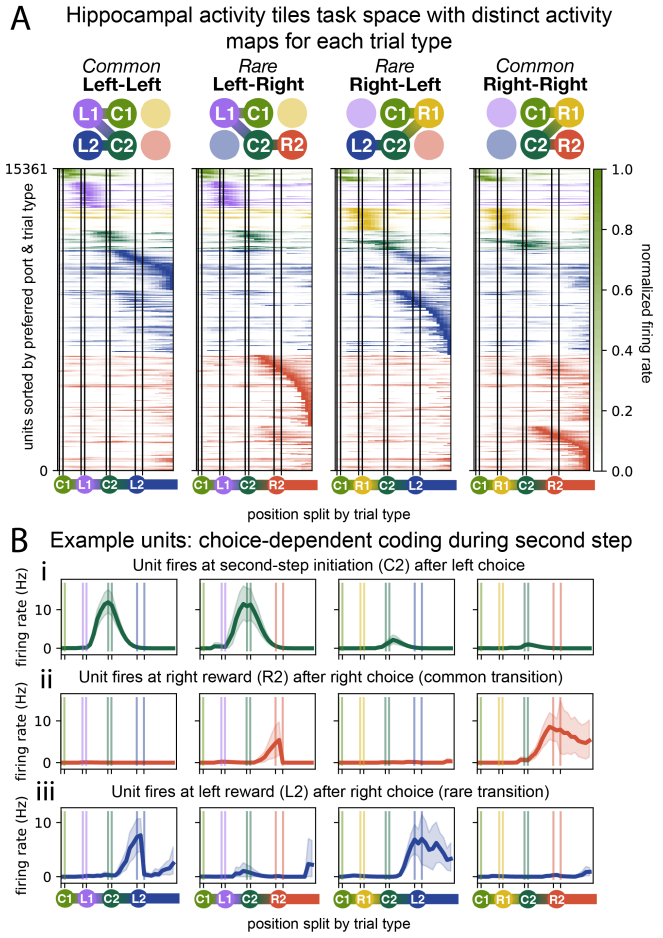

Figure 2: Hippocampal activity encodes task states and choice history.

Choice encoding between trials is biased towards upcoming, model-based choices

Choice encoding during the ITI shows a model-based dependence on reward and transition. To understand how the hippocampus support planning, we measure neural encoding during the inter-trial interval (ITI): if the animal links actions and outcomes to plan before selecting the next action, this must occur after an outcome is observed and before the next action takes place. Supporting this notion, the ITI during the two-step task has been identified as a critical period for learning in the OFC to occur (Miller, Botvinick, & Brody, 2022). We constrain analysis to two time windows in the ITI: 1 second following outcome port entry, which we refer to as **post-trial**; and 1 second preceding initiation port entry of the next trial, which we refer to as **pre-initiation**. This ensures that the rat remains near the outcome port post-trial or is moving towards the initiation port pre-initiation, minimizing the risk of measuring while the rat is disengaged from the task.

Having choice dependent encoding throughout the trial allows us to use a Bayesian reconstruction approach (Zhang, Ginzburg, McNaughton, & Sejnowski, 1998) to determine which choice is most likely being encoded during the ITI. We look specifically at the relative likelihood of choice, where positive values correspond to encoding of the action chosen prior to the ITI, and negative values correspond to encoding of the non-chosen action. We find that choice encoding within both ITI windows is biased by both reward and transition; specifically, rats are more likely to encode the chosen action following a common-transition reward and a rare-transition omission, and are more likely to encode the non-chosen action following a common-transition omission and a rare-transition reward (Figure 3A). This pattern of choice encoding is consistent with a MB dependence on reward and transition, where common-transition reward and rare-transition omissions lead to repeating actions, and common-transition omission and rare-transition rewards lead to switching actions.

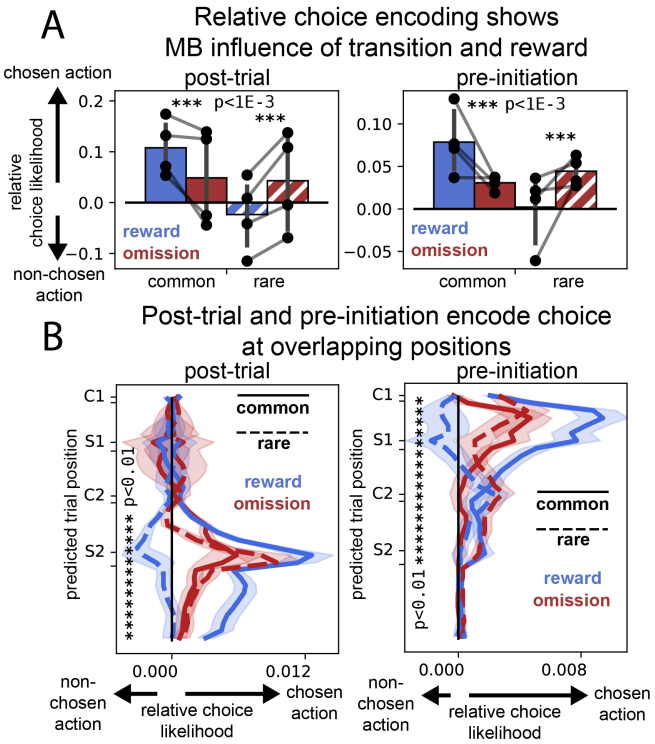

Figure 3: Relative choice encoding during the inter-trial interval.

Positions updated post-trial are retrieved pre-initiation. Planning requires that the value learned at some outcome is propagated to relevant choices before another action is selected. We use a similar reconstruction approach to not only decode relative choice, but also the position being encoded

in both ITI windows (Figure 3B). We find that post-trial encoding is biased towards second-step positions and reward (after C2), and pre-trial encoding is biased towards first-step positions (before C2). However, there exists a window of overlap in encoded positions between C2 and S2. Importantly, the same MB choice information is being encoded at these overlapping positions, suggesting that choice information encoded post-trial is being retrieved pre-initiation to be associated with upcoming choice states.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1656466, a Research Innovation Award from Princeton Neuroscience Institute, and NIMH R01MH121093.

References

Duvelle, É., Grieves, R. M., & van der Meer, M. A. (2023, January). Temporal context and latent state inference in the hippocampal splitter signal. *eLife*, *12*, e82357. doi: 10.7554/eLife.82357

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

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2022, August). Value representations in the rodent orbitofrontal cortex drive learning, not choice. *eLife*, *11*, e64575. doi: 10.7554/eLife.64575

Zhang, K., Ginzburg, I., McNaughton, B. L., & Sejnowski, T. J. (1998, February). Interpreting Neuronal Population Activity by Reconstruction: Unified Framework With Application to Hippocampal Place Cells. *Journal of Neurophysiology*, *79*(2), 1017–1044. doi: 10.1152/jn.1998.79.2.1017