Closed-coop prioritized hippocampal replay in brain-machine interface of value-biased choice

William Yixiang Zheng

https://youtu.be/M463StuUfFo

Abstract

Sharp-wave ripples (SWRs) and hippocampal neural sequences are crucial for spatial navigation and memory. In artificial reinforcement learning (RL), prioritized experience replay accelerates learning by oversampling high-value experiences from the replay buffer. However, whether hippocampal replay is systematically organized by choice value or can be causally biased to accelerate learning remains unknown. Here we propose a hippocampal brain-machine interface (BMI) experiment while rats will perform a probabilistic virtual-reality T-maze with asymmetric rewards, with higher-reward arm pseudo-randomly assigned across sessions. We aim to first decode place cell sequences within individual theta cycles and SWR events to quantify value-dependent biases toward the higher-reward arm and assess whether these biases predict trial-by-trial choices. Then, we aim to implement a closed-loop "prioritized replay" by delivering brief optogenetic stimulation when SWR replay represents the high-reward arm, and compare learning speed and sequence metrics to control sessions. I predict that selectively reinforcing high-value replay will increase replay priority, strengthen prospective sweep bias, and accelerate acquisition of the optimal route. This work will inform the mechanisms underlying episodic memory and imagination.

Introduction

Animals constantly explore their environment and evaluate surroundings to navigate toward higher-reward choices. For example, when you move to a new city for work, you may initially sample several coffee shops, but over time you learn which one offers the best experience and begin going there most often while only occasionally exploring the others. These behaviors imply that the brain not only represents potential choices, but also assign and update their expected values. Different neural mechanisms have been proposed to support such value-biased choice and decision making, with hippocampal mechanisms playing a central role [1].

The hippocampus is crucial for episodic memory and spatial navigation, supporting internal cognitive maps that allow animals to represent spatial locations, track trajectories, and plan routes toward goals [2]. Specifically, hippocampal place cells fire at the specific location to support the formation of internal cognitive maps at the population level [3]. Beyond coding static locations, place-cell activity is structured by the theta rhythm into temporally compressed sequences that start slightly behind the animal and extend forward along potential routes ("theta sweeps") [4]. During pauses in movement and sleep, related sequences are expressed in further compressed form during sharp-wave ripple (SWR) events, replaying recent or prospective trajectories [5]. These internally generated sequences are natural substrates for model-based evaluation and credit assignment, providing a mechanism for hippocampal-cortical network to simulate potential trajectories, estimate their expected value, and update memories for behaviorally important routes [6].

In artificial reinforcement learning (RL), experience replay helps agents learn from stored interactions (state, action, reward, and next stage), and "prioritized experience replay" accelerates learning by oversampling experiences that are more valuable or informative [7]. This

suggests that hippocampal theta sweeps and SWR replay may implement a biological form of prioritized replay. However, it remains unknown whether hippocampal sequences are systematically biased by value in this way or whether selectively enhancing high-value replay can causally accelerate learning. Emerging hippocampal brain-machine interfaces (BMIs) that decode population activity in real time and deliver closed-loop interventions now make this question experimentally tractable [8]. In this mini-grant, I will test the hypothesis that hippocampal theta sweeps and SWR replay preferentially represent higher-value trajectories, and that a hippocampal BMI that selectively reinforces these events will strengthen this bias and accelerate learning of rewarded routes.

Specific Aims

Aim 1: Quantify value-biased hippocampal sequences that predict future choices in a probabilistic T-maze

We will test whether internal hippocampal sequences are biased toward higher-value options and whether this bias predicts choice. We will use the head-fixed rats performing a virtual-reality T-maze in which the two arms deliver unequal reward probabilities, with the identity of the high-reward arm (left vs. right) pseudo-randomly counterbalanced across sessions. We will record dorsal CA1 population activity with high-density probes and train a neural decoder to map ensemble spiking as position and trajectory identity (left vs. right arm) in real time [9]. Using this decoder, we will quantify value bias in theta sweeps at the choice point and in SWR replay during rest, and test whether these sequence-based indices predict trial-by-trial choices with the control of reward history and sensorimotor variables (e.g. running speed and licking position).

Aim 2: Implement a closed-loop "prioritized replay" BMI that selectively reinforces high-value replay and test its impact on learning

Building on Aim 1, we will ask whether manipulating replay content can causally bias hippocampal sequences and accelerate learning. We will implement a closed-loop hippocampal BMI that detects SWRs and decodes replay content online [8] [10]. When replay sequences represent the currently high-value arm, a brief optogenetic stimulation will be delivered to enhance synaptic plasticity associated with those replayed trajectories [11]. We will compare behavioral learning speed, choice performance, and hippocampal sequence metrics (replay priority: relative replay frequency of high- vs. low-value arms, and theta sweep bias at the choice point) between prioritized-replay stimulation sessions and control conditions in which stimulation is absent or not contingent on replay content [12].

Research Design

After recovered from surgery, rats will be handled daily and accommodated to the experimenter and recording. One day before training, rats will be placed under water restriction. Rats will be trained to navigate a virtual-reality T-maze (50 cm central arm and 50 cm left/right arms). Within each session, the high-reward arm will remain fixed so that animals can learn a stable mapping between arm and reward probability, enabling interpretable value-biased theta sweeps and replay. Across sessions, the identity of the high-reward arm (left vs. right) will be counterbalanced in a pseudo-random order to dissociate spatial side from value and avoid confounding variables. Animals will perform daily sessions with sufficient trials to track how behavior changes as they discover the higher-probability arm. A trail will be completed once the animal licks at the water port on the chosen arm and returns to the starting point of the central arm.

Neural activity will be recorded from dorsal CA1 using high-density probes [13]. A deep neural decoder will be trained on spiking data during overt navigation to map population activity onto 2D position and arm identity (left vs. right). Once trained, the decoder will be used both offline and in low-latency online mode to estimate the trajectory content from ensemble activity [8].

For theta sweeps, we will focus on a pre-choice window on the central arm (1-2 seconds before the choice point). Theta cycles will be identified in the local field potential (LFP) and spikes will be segmented by cycle. Within each cycle, we will use the decoder to reconstruct temporally compressed trajectories and classify each theta sequence as preferentially representing the left or right arm, discarding low-confidence events [14]. To quantify value bias, we will compute a Sweep Bias Index (SBI) defined as

$$\mathrm{SBI} = \frac{N_{sweeps\ to\ high-reward\ arm} - N_{sweeps\ to\ low-reward\ arm}}{N_{sweeps\ to\ high-reward\ arm} + N_{sweeps\ to\ low-reward\ arm}}$$

We will track this index over trials and across learning, and use logistic regression to test whether trial-by-trial variation in sweep counts and SBI predicts upcoming choice, while controlling for head direction and recent rewards [15].

To assess value bias in replay, we will analyze SWRs during brief pauses and inter-trial intervals. We will detect SWRs in the LFP and decode position and arm identity over time to identify replayed trajectories, and classify events as representing the high- or low-reward arm [16]. A Replay Priority Index (RPI), defined analogously to SBI as

$$\mathrm{RPI} = \frac{N_{replays\ of\ high-reward\ arm} - N_{replays\ of\ low-reward\ arm}}{N_{replays\ of\ high-reward\ arm} + N_{replays\ of\ low-reward\ arm}}$$

will quantify the degree to which replay favors the higher-value trajectory. We will examine how RPI evolves as animals learn the task and whether it covaries with behavioral performance and SBI [16]. Together, these analyses will test whether internal hippocampal sequences exhibit a systematic value bias and whether that bias prospectively predicts choice.

To implement a closed-loop "prioritized replay" BMI that selectively reinforces high-value replay, we will extend the decoder and replay detection pipeline to operate online [8]. During rest periods and inter-trial intervals, SWRs will be detected in real time from the LFP, and CA1 population activity during each event will be passed through the trained decoder to classify replayed trajectories as high-reward arm, low-reward arm, or ambiguous [17]. In "prioritized replay" sessions (Stim), when a SWR is decoded as representing the currently higher-reward arm, a brief optogenetic stimulation train will be delivered with minimal latency using a 470-nm LED (Prizmatrix). Replays of the low-reward arm will not trigger stimulation (No-Stim). Stimulation will target either intrinsic hippocampal circuits (e.g. CA3-CA1) or dopaminergic/neuromodulatory input (e.g. VTA terminals in hippocampus) to enhance plasticity associated with high-value replay [18]. Control conditions will include No-Stim and Random-Stim sessions.

We will analyze the impact of prioritized replay stimulation at both behavioral and neural levels. Behaviorally, we will compare learning speed (e.g. trials to reach $\geqslant$ 80% choice of the high-reward arm within a session). We will fit the RL models to individual animals' choices to estimate learning rates and decision noise, asking whether prioritized replay stimulation selectively increases effective learning rates relative to the control group. In neural aspect, we

will recompute the RPI under Stim and Control conditions to test whether closed-loop stimulation increases the proportion and fidelity of SWR events representing the high-reward arm. Replay Fidelity (RF) will be quantified by correlating time-resolved population vectors during each replay with place-field templates obtained from behavior, separately for each arm and position [19]. We predict higher fidelity for high-reward replay in the Stim condition. Finally, we will examine theta sweeps at the choice point following periods of prioritized replay stimulation (Stim), testing whether SBI toward the high-reward arm is enhanced, consistent with replay-driven shaping of prospective planning.

If hippocampal sequences indeed support a prioritized replay strategy, we expect that high-reward trajectories will be overrepresented in theta sweeps and SWR replay, and that selectively reinforcing those events will further increase this priority, strengthen prospective sweep bias, and measurably accelerate learning of rewarded routes.

Limitations

In this work, we expect to provide causal evidence that value-biased hippocampal replay contributes to learning and that selectively reinforcing high-value sequences can shape planning. However, the mechanisms underlying closed-loop optogenetic effects may be difficult to interpret. Even if prioritized replay stimulation increases replay bias and speeds learning, these effects could arise from nonspecific changes in arousal and motivation. To address this, control conditions (No-Stim and Random-Stim session), basic behavioral variables (e.g. running speed and licking rate), and single-unit analyses of place-cell stability and remapping are required to ensure that observed changes in sequence content are not driven by global state changes or recording instability.

Besides, the behavioral and recording context limits generalization. A head-fixed virtual T-maze with binary probabilistic choices captures only a narrow subset of naturalistic spatial navigation, and hippocampal dynamics may differ in freely moving animals or in richer decision spaces. This experiment also only focuses on dorsal CA1, leaving potential contributions of other hippocampal subfields and interconnected regions (e.g. prefrontal cortex) unmeasured. Thus, future studies need to be done across multiple brain regions with more complex, natural behavioral tasks.

Conclusion

In this project, I will test whether hippocampal theta sweeps and SWR replay implement a biological form of prioritized experience replay inspired by RL, and whether selectively reinforcing high-value replay can causally accelerate learning and bias planning. By combining a probabilistic virtual-reality T-maze, high -density CA1 recordings with real-time decoding, and closed-loop optogenetic stimulation, I will quantity value biases in hippocampal sequences and manipulate them with a hippocampal BMI to uncover the neural implementation of an internal world model.

This work will provide mechanistic evidence that internal hippocampal sequences are not passive replays the experiences, but are actively prioritized toward high-value choices and trajectories, and can be used to bias future decisions. Beyond uncovering the mechanism of learning and planning, this work will help bridge concepts from artificial reinforcement learning and biological memory systems, and inform the design of neuro-inspired BMIs that leverage internal models to enhance learning and flexible, goal-directed behavior.

References

1. A. Bakkour et al., *eLife* 8:e46080 (2019).
2. N. Nyberg et al., *Neuron* 110, 394-422 (2021).
3. J. Ormond, J. O'Keefe, *Nature* 607, 741-746 (2022).
4. N. Burgess et al., *Current Opinion in Neurobiology* 21 (5), 734-744 (2011).
5. D. K. Roumis et al., *Current Opinion in Neurobiology* 35, 6-12 (2015).
6. J. J. W. Bakermans et al., *Nature Neuroscience* 28, 1061-1072 (2025).
7. H. Li et al., *Scientific Reports* 14, 6014 (2024).
8. C. Lai et al., *Science* 382, 566-573 (2023).
9. J. Bono et al., *eLife* 12:e80671 (2023)
10. W. Fang, A. G. Mombeini et al., *Current Opinion in Behavior Sciences* 66, 101597 (2025)
11. L. Z. Fan, D. K. Kim, J. H. Jennings et al., *Cell* 186 (3), 543-559.e19 (2023)
12. Y. Liu et al., *Science* 372, 807 (2021).
13. Z. Ye et al., *Neuron* 113 (23), 3966-3982.e12 (2025)
14. E. Parra-Barrero et al., *eLife* 10:e70296 (2021)
15. A. Z. Vollan et al., *Nature* 639, 995-1005 (2025)
16. E. L. Krause et al., *Neuron* 110 (4), 722-733.e8 (2022)
17. A. K. Gillespie et al., *Neuron* 109 (19), 3149-3163.e6 (2021)
18. Z. Brzosko et al., *Neuron* 103 (4), 563-581 (2019)
19. H. Xu et al., *Neuron* 101 (1), 119-132.e4 (2019)

Acknowledgement

I would like to thank Dr. Sun and and Ivan Kondratyev for supervising me on this course. The YouTube like is here: https://youtu.be/M463StuUfFo

If you have any question please contact ywz2@cornell.edu