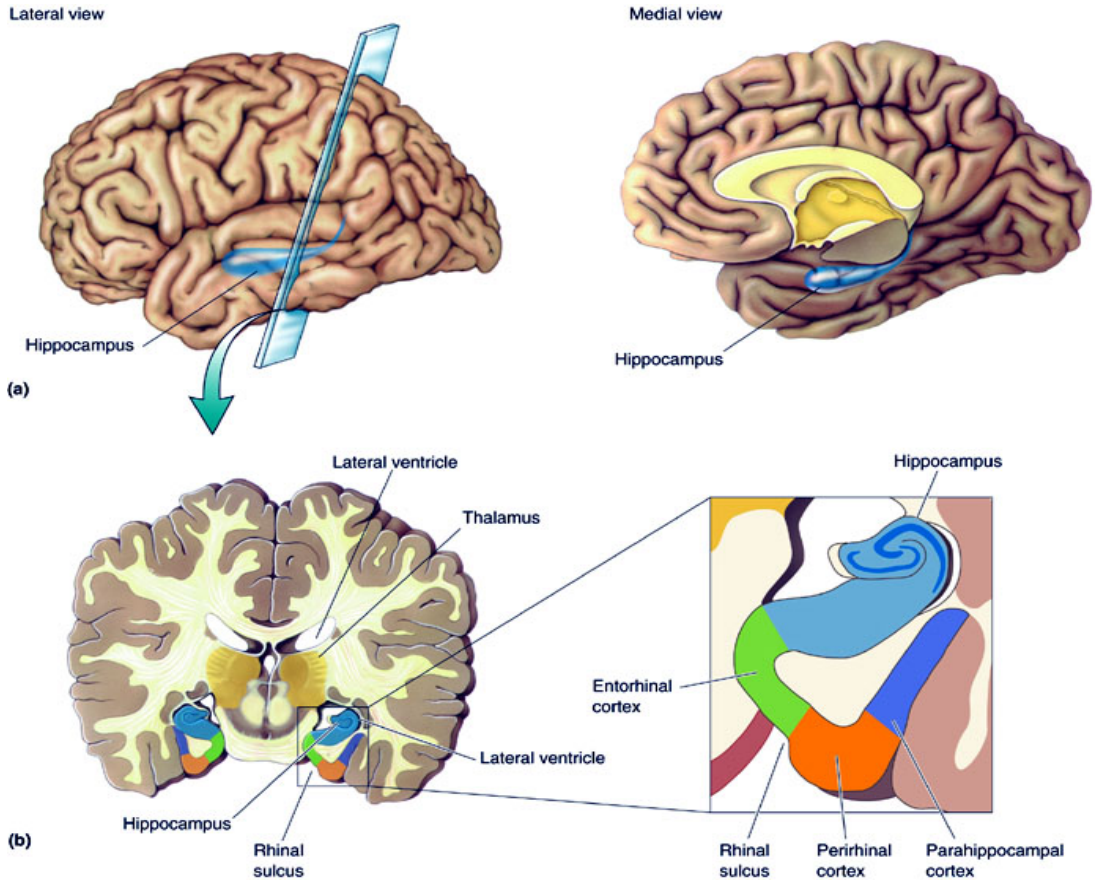
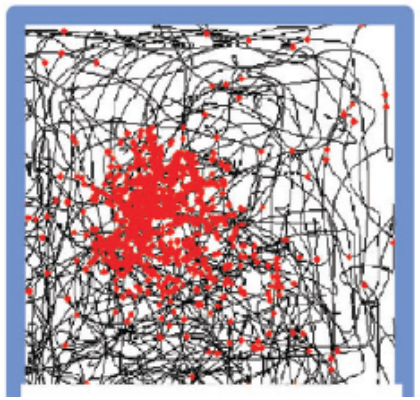
Emergent Replay Coordinates Context and Static Representation in Artificial Agents

Hippocampus as a memory and navigation machine

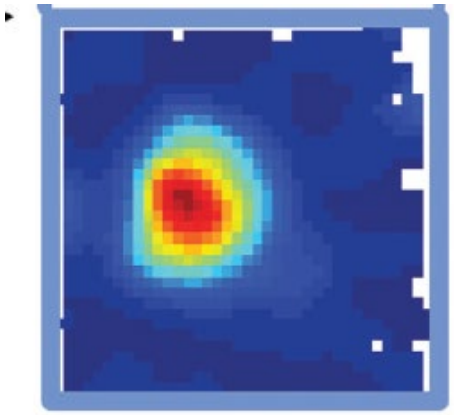


Bear, M., Connors, B., & Paradiso, M. A. (2020). *Neuroscience: exploring the brain, enhanced edition: exploring the brain*. Jones & Bartlett Learning.

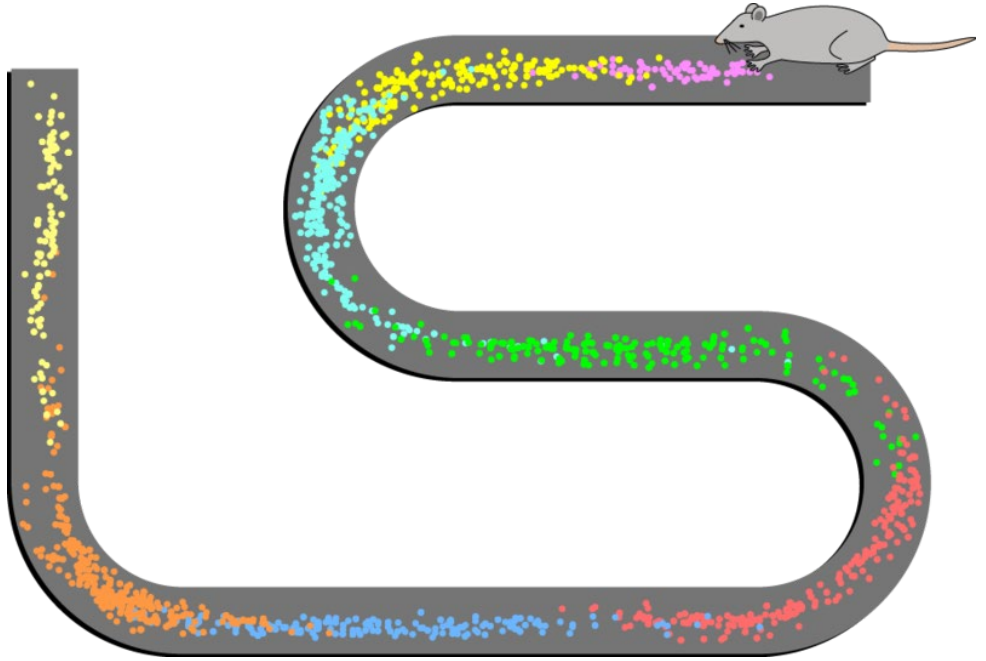
Roles in Navigation



Trajectory&
Place cell's
firing rate



Place field

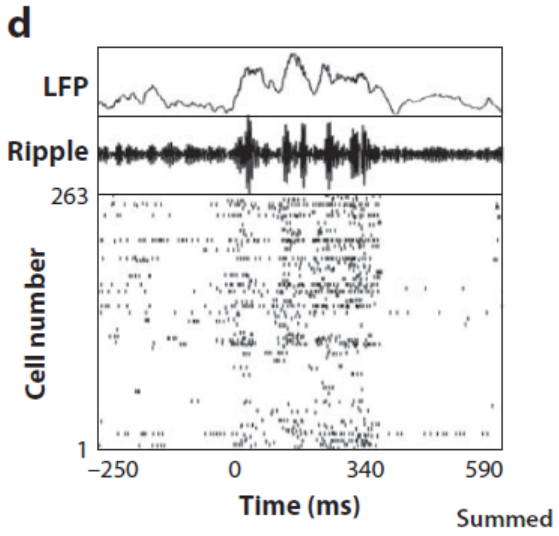


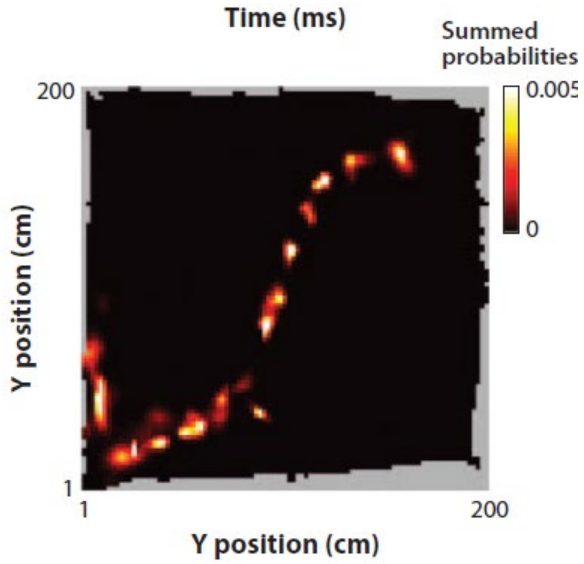
The place cells constitute **a cognitive map** through which navigation is possible and efficient.

Moser, M. B., Rowland, D. C., & Moser, E. I. (2015). Place cells, grid cells, and memory. *Cold Spring Harbor perspectives in biology*, *7*(2), a021808.

By Stuartlayton at English Wikipedia, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=43746578

Replay





Definition:

Reactivation of a sequence pattern of place cells that once fired during movement

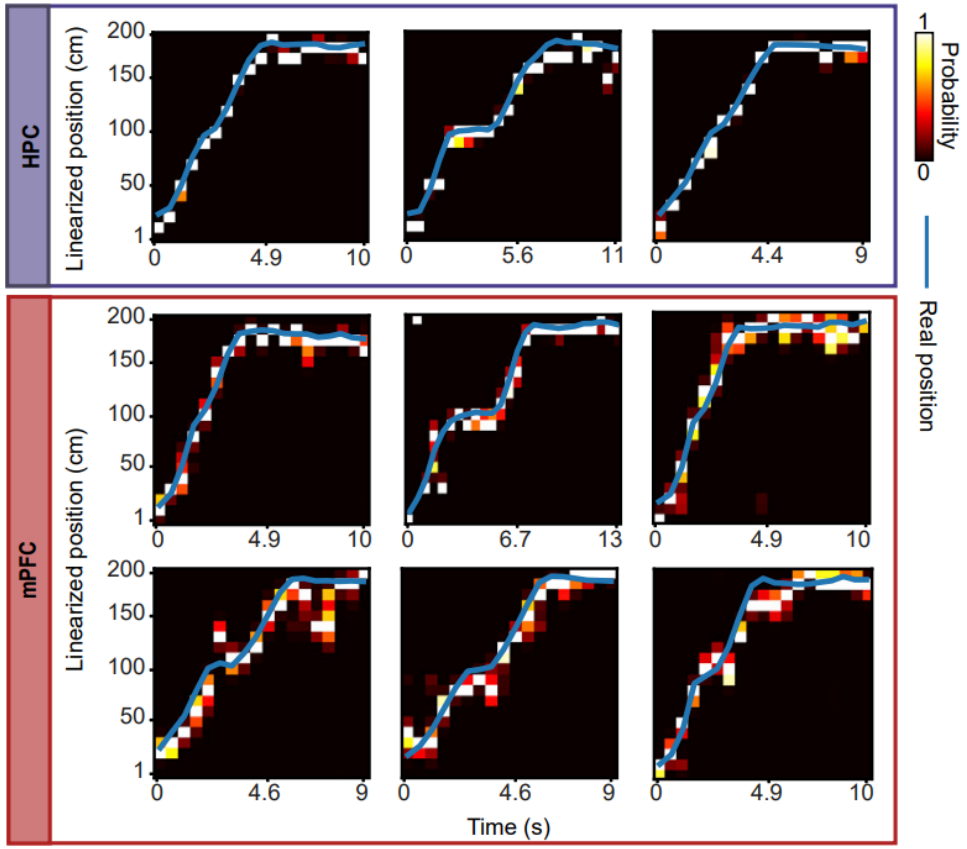
"Reexperiencing"/"Organizing"

Putative functions

- **Memory consolidation/ Memory retrieval/ Spatial memory** (overlap with previous taken path)
- **Planning/ Navigation** (overlap with future taken path)

Foster, D. J. (2017). Replay comes of age. *Annual review of neuroscience*, *40*, 581-602.

Replay also exists in cortex



Cortex replays are often coordinated with hippocampal replays.

Further support memory consolidation and navigation…

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., & Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature neuroscience*, *12*(7), 919-926.

Current Research

(1)Theories relies heavily on their hypotheses , which have been complex and less convincing.

(2)Experimental results are solid themselves, but they often support different functions and contradict each other.

Our Goal and Method

Let biological representations emerge as a natural result of task-optimized (goal-directed) training. Like Garbor in the vision.

Deep reinforcement learning (DRL) is employed with the minimal sufficient conditions:

(1) Replay serves for reward maximization. (we know block of SWR (sharp wave-ripple) causes significant harm to learning performance)

(2) Replay is a form of communication between the neocortex and hippocampus.

Our Goal and Method

To test whether these two conditions are sufficient, our model consists of two main sub-modules: hippocampus (HPC) module and medial prefrontal (mPFC) module.

To test whether these two conditions are sufficient, our model consists of two sub-modules: hippocampus(HPC) module and medial prefrontal(mPFC) module.

1. The HPC serves as a cognitive map (path integration), and as a short-term memory (episodic memory).

2. The mPFC gathers information from the sensory cortex and HPC to make decisions.

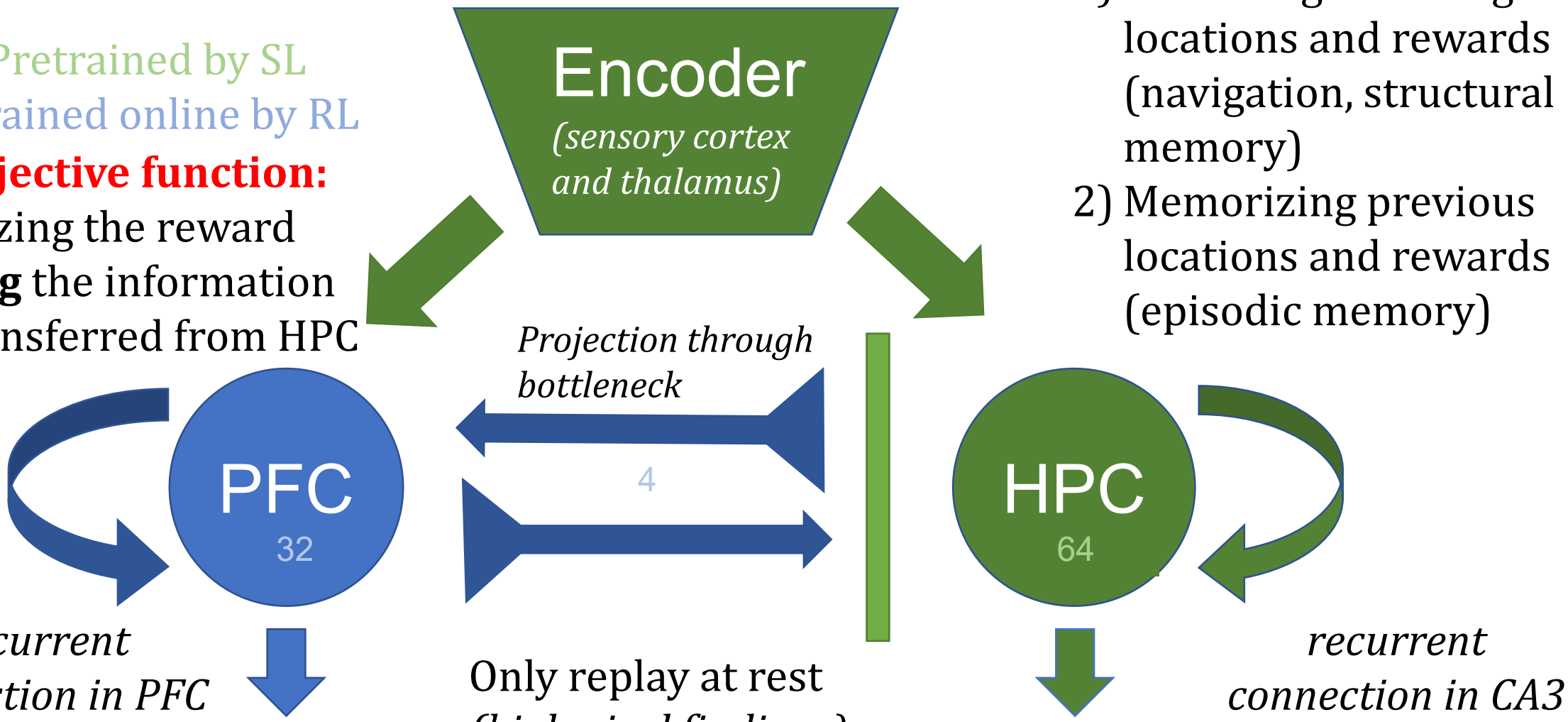
We do not assume in advance what kind of information should be conveyed from HPC to PFC !

Model Structure

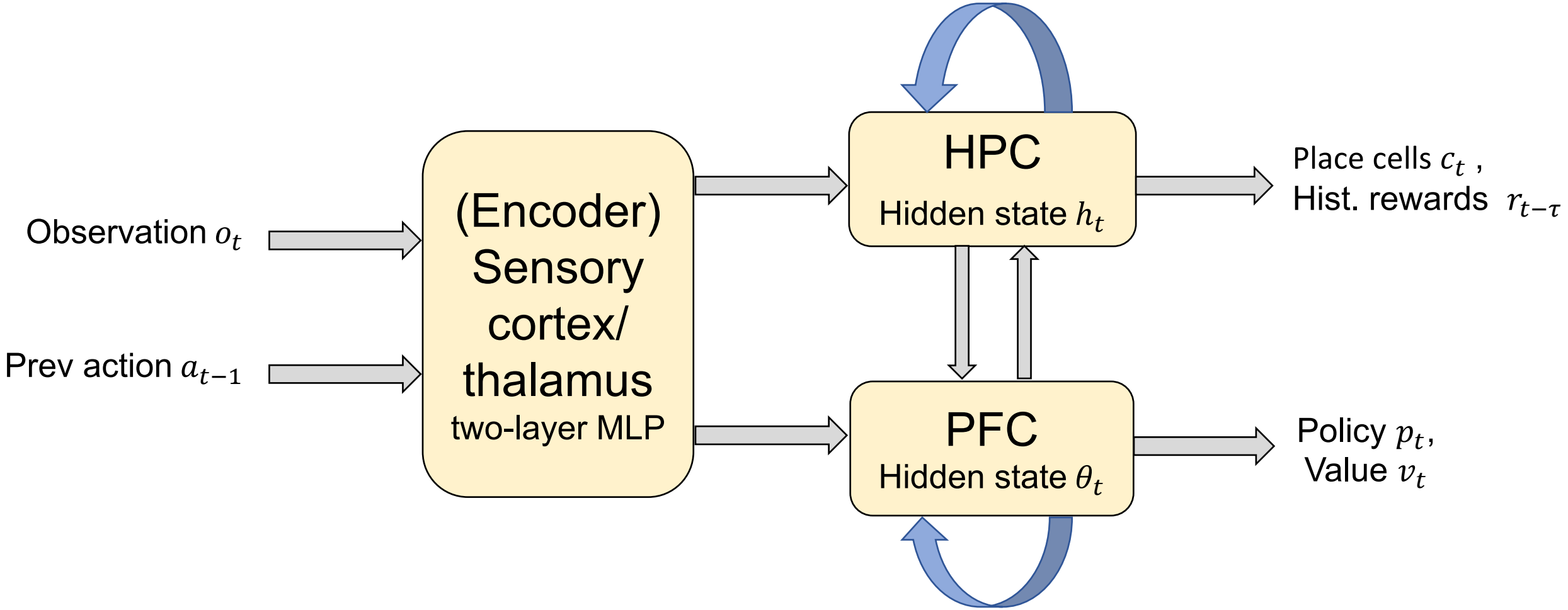
Green: Pretrained by SL
Blue: Trained online by RL

PFC Objective function:
Maximizing the reward **utilizing** the information flow transferred from HPC

HPC Objective function:
1) Predicting following locations and rewards (navigation, structural memory)
2) Memorizing previous locations and rewards (episodic memory)

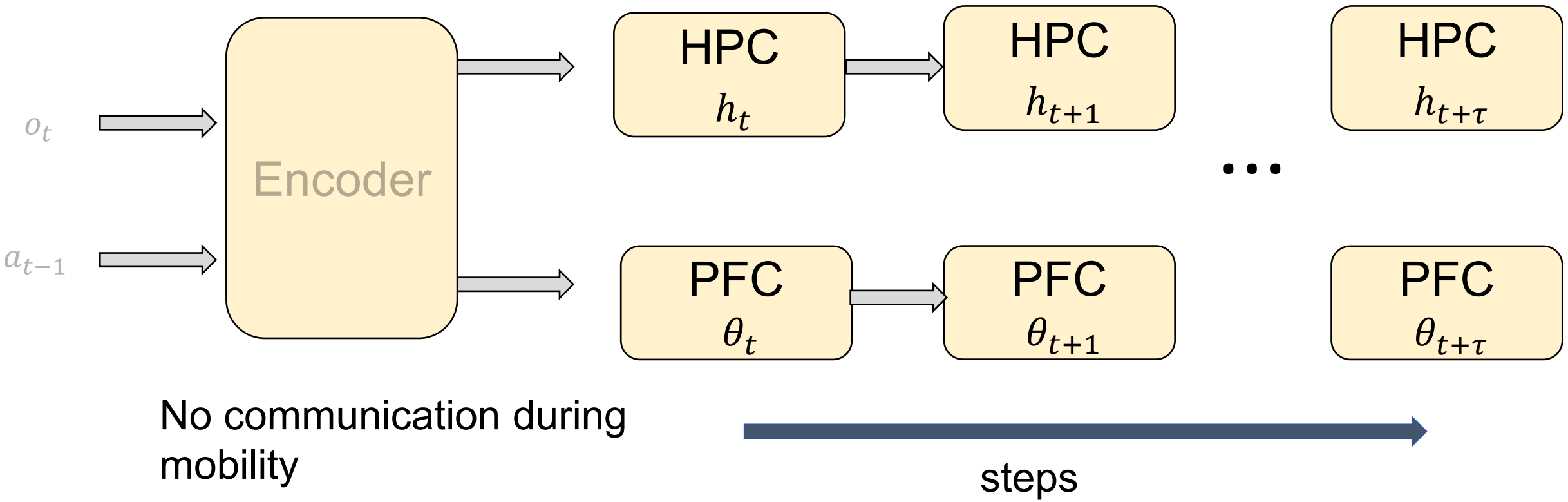


Model Structure (Inputs and Outputs)

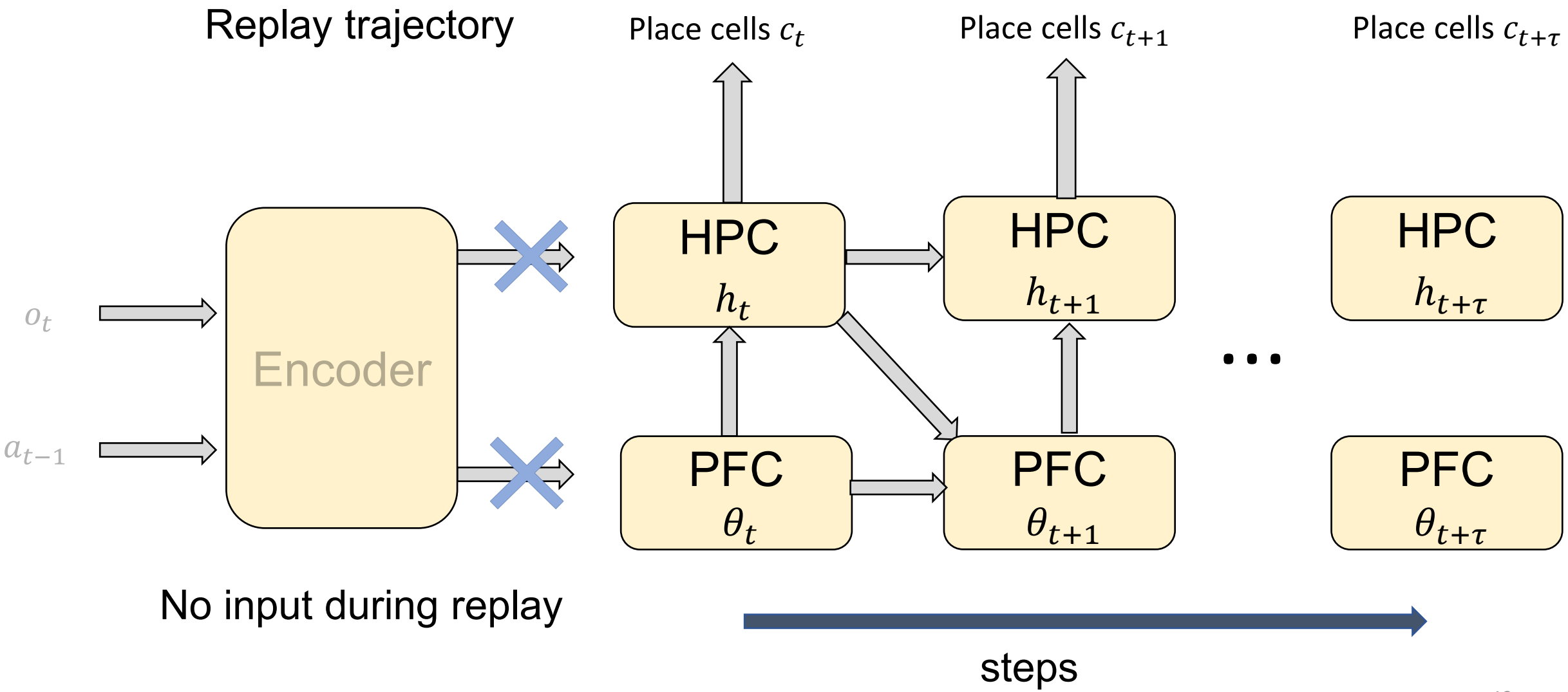


An Action is sampled based on probabilities of different options in Policy.

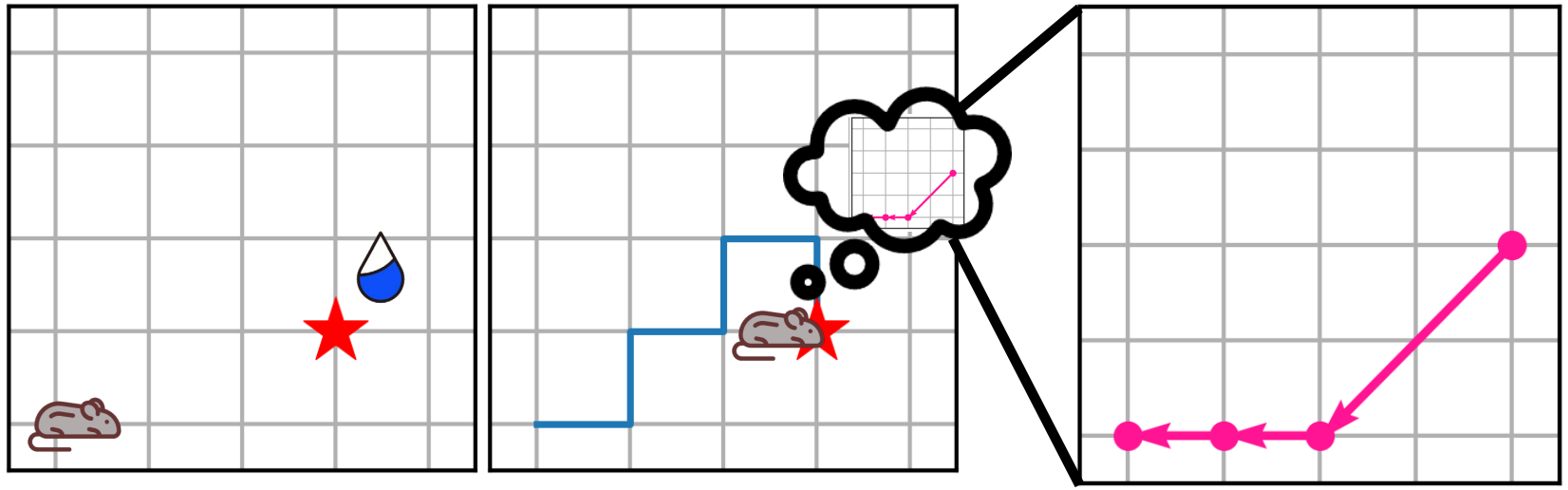
How Modules Work during Mobility

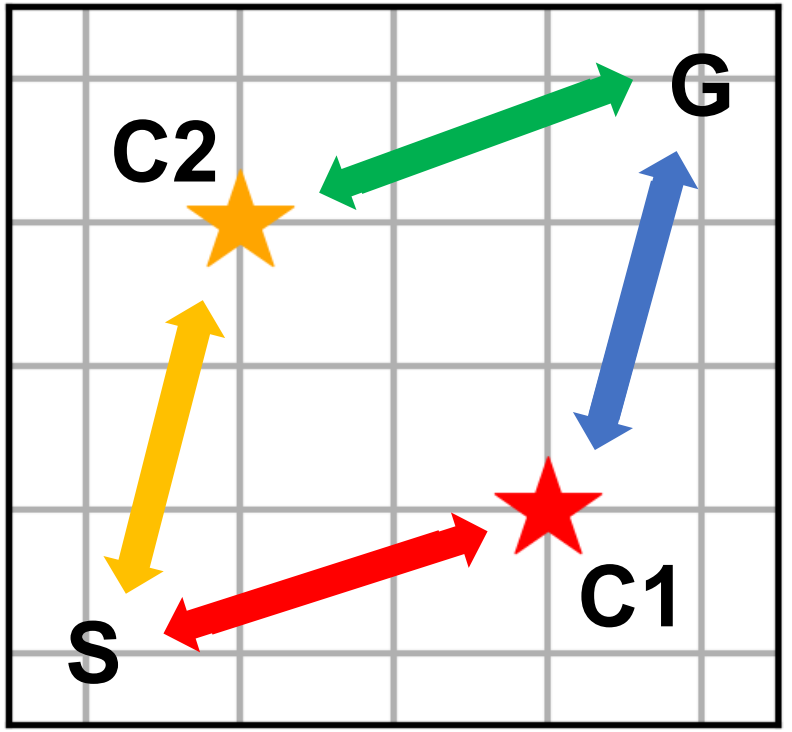


How Modules Work during Replay



Experiments (Mice)

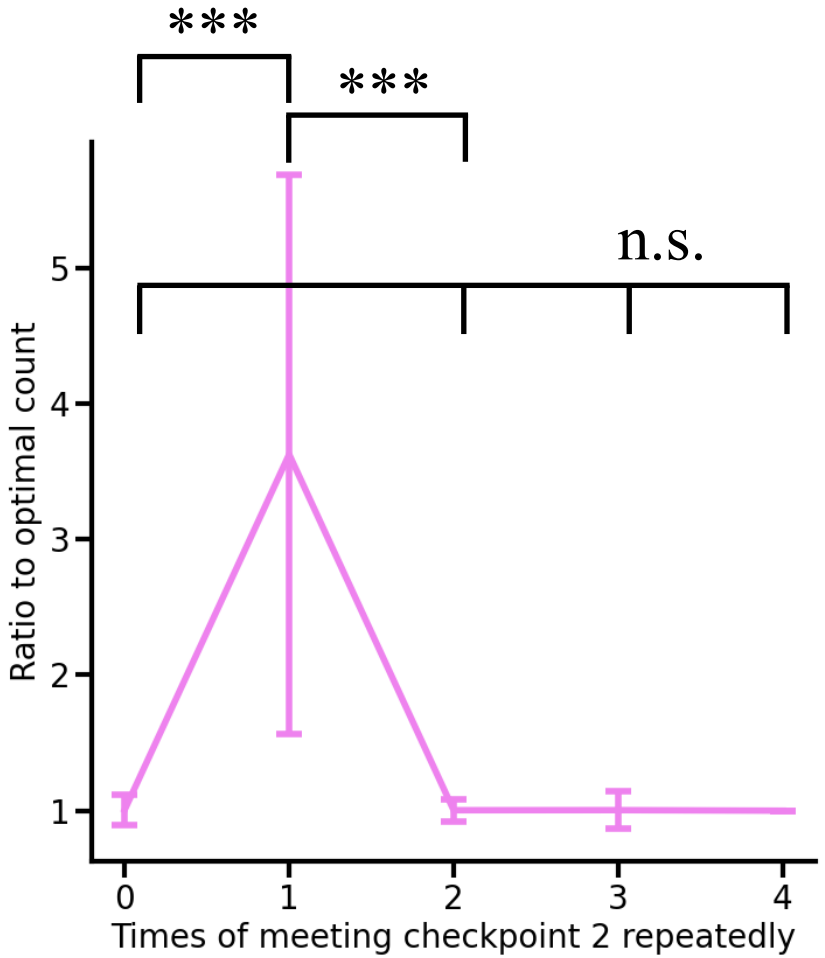




The blue line represents the replay trajectory;
The red line represents last replay trajectory when it reaches the goal.
The number of replay steps are set to be 4 after 1 step of action.

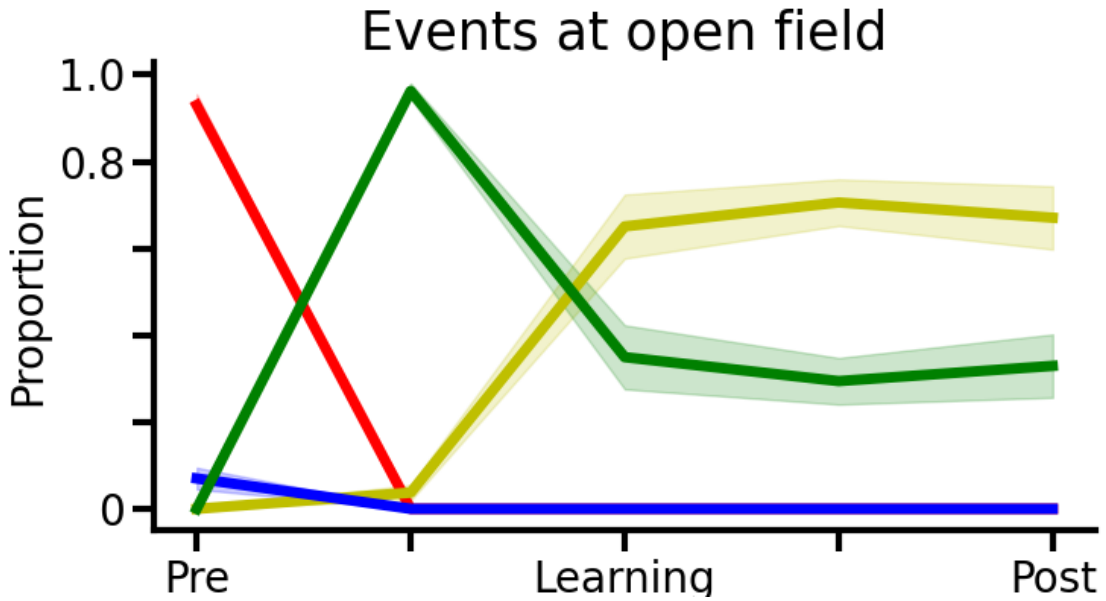
The mouse started from S, obtained a small reward at C1, and then reached G getting a big reward. We investigate changes in behavior and replayed trajectories when the reward switched from C1 to C2.

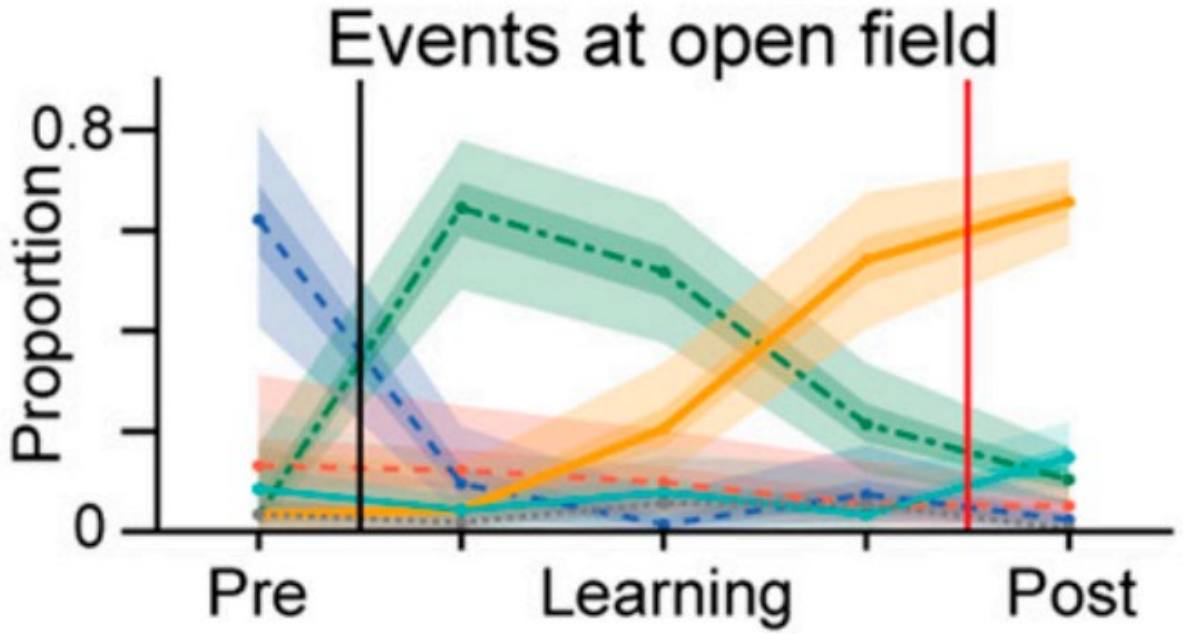
Experiments



Performance of the AI agent when it comes across the new checkpoint, represented by the ratio of the number of agents' average steps to the optimal trajectory (in this task, the optimal number of steps is 3).

Experiments

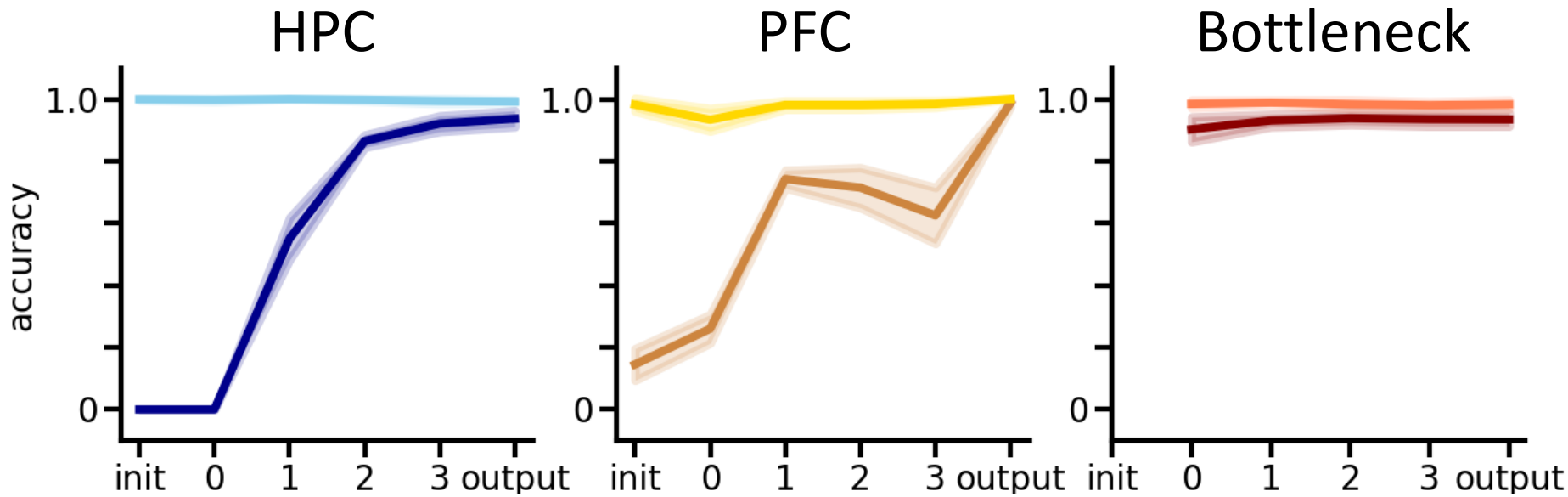




Proportion of replay trajectories representing different segments of the room.
Left: our distribution of segment replay as learning progresses.
Right: experimental results from Igata et al. 2021
High similarity in two aspects. First, the amount of replay of S-C2 increases.
Second, the amount of C2-G first increases and then drops down.

Experiments

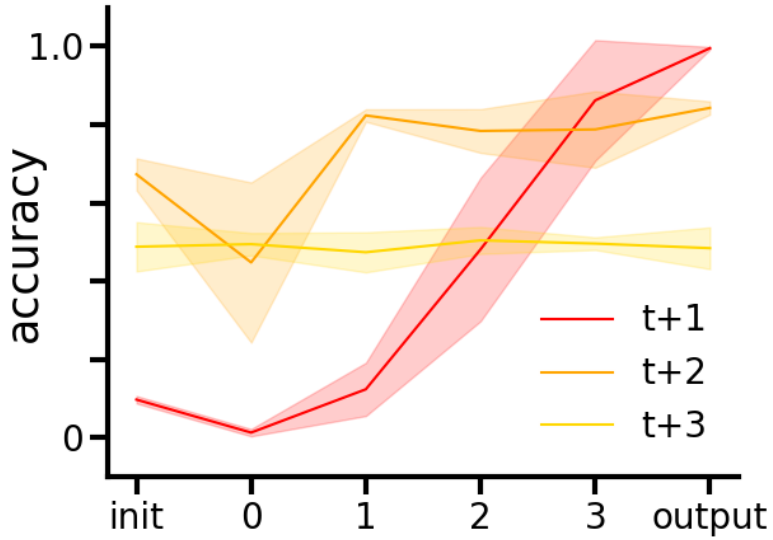
Phase decoder



Decoding accuracy for the correct checkpoint location of the activities in HPC, PFC and the information passage at different steps
(x axis, initial: before replay, 1-4: different steps during replay, output: after replay).
Light color: the first time the agent meets the new checkpoint;
Dark color: the second time.

Experiments

Action decoder



Decoding accuracy for the correct actions in the following process of the activities in HPC, PFC and the information passage at different steps.
(x axis, initial: before replay, 1-4: different steps during replay, output: after replay).
This shows future planning ability.

Experiments

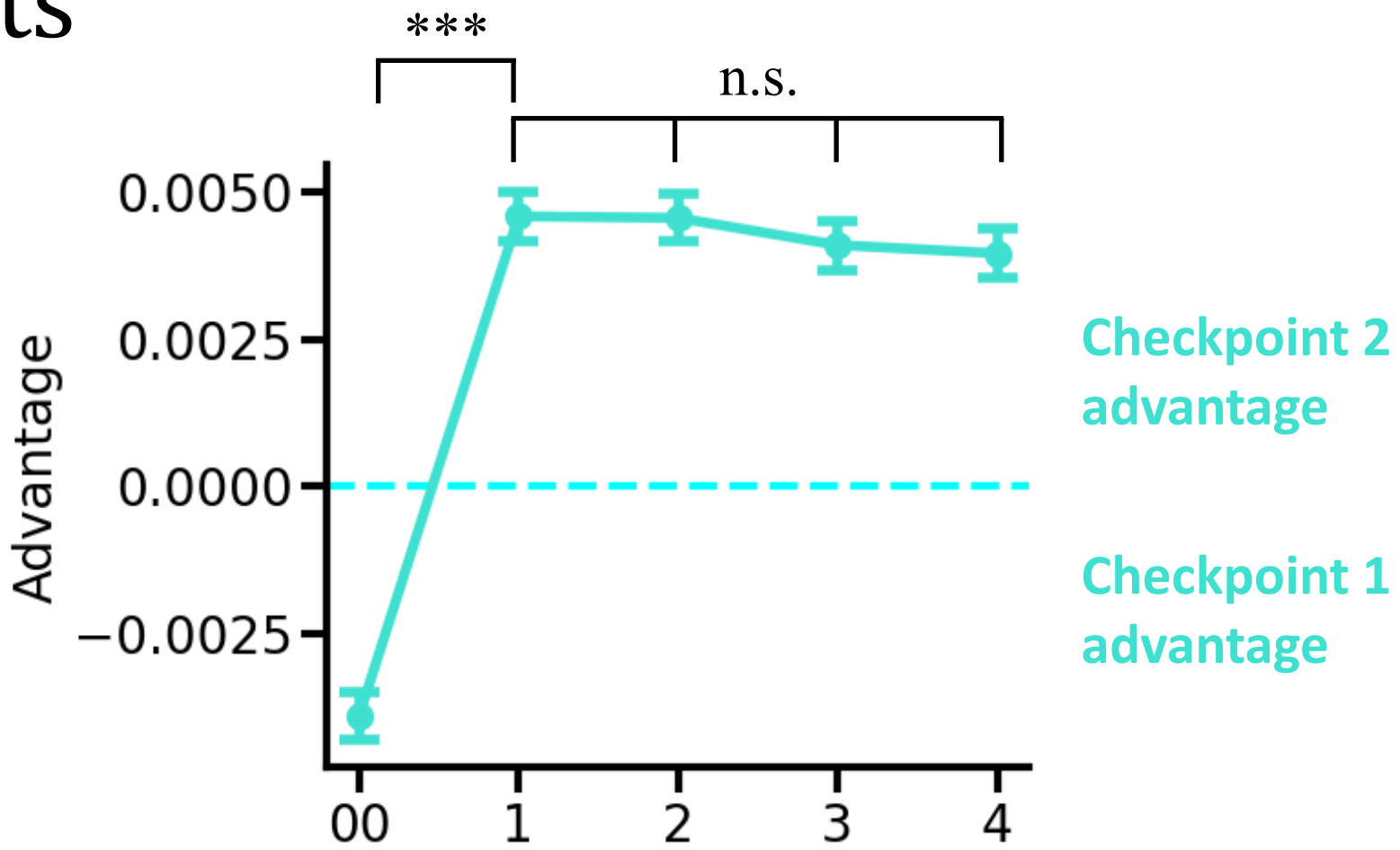
Value map

Before change | 1st Time | 2nd Time | 3rd Time | 4th Time

Value map as a result of "stop and scan" when the agent meets the checkpoint. The leftmost is the value map before the checkpoint changes (from C1 to C2), and the others corresponds to different times the agent meets the same checkpoint repeatedly. Notice the most obvious change occurs at the location of the checkpoint.

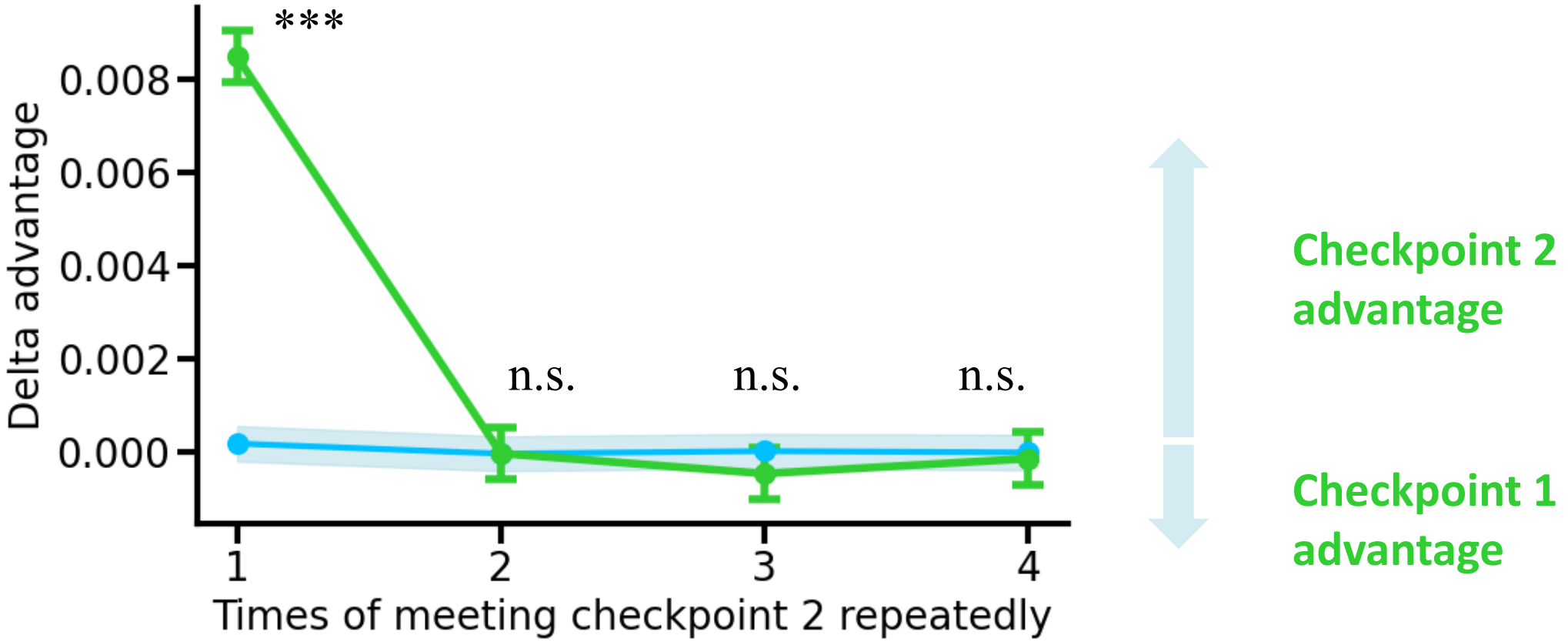
Experiments

Value advantage



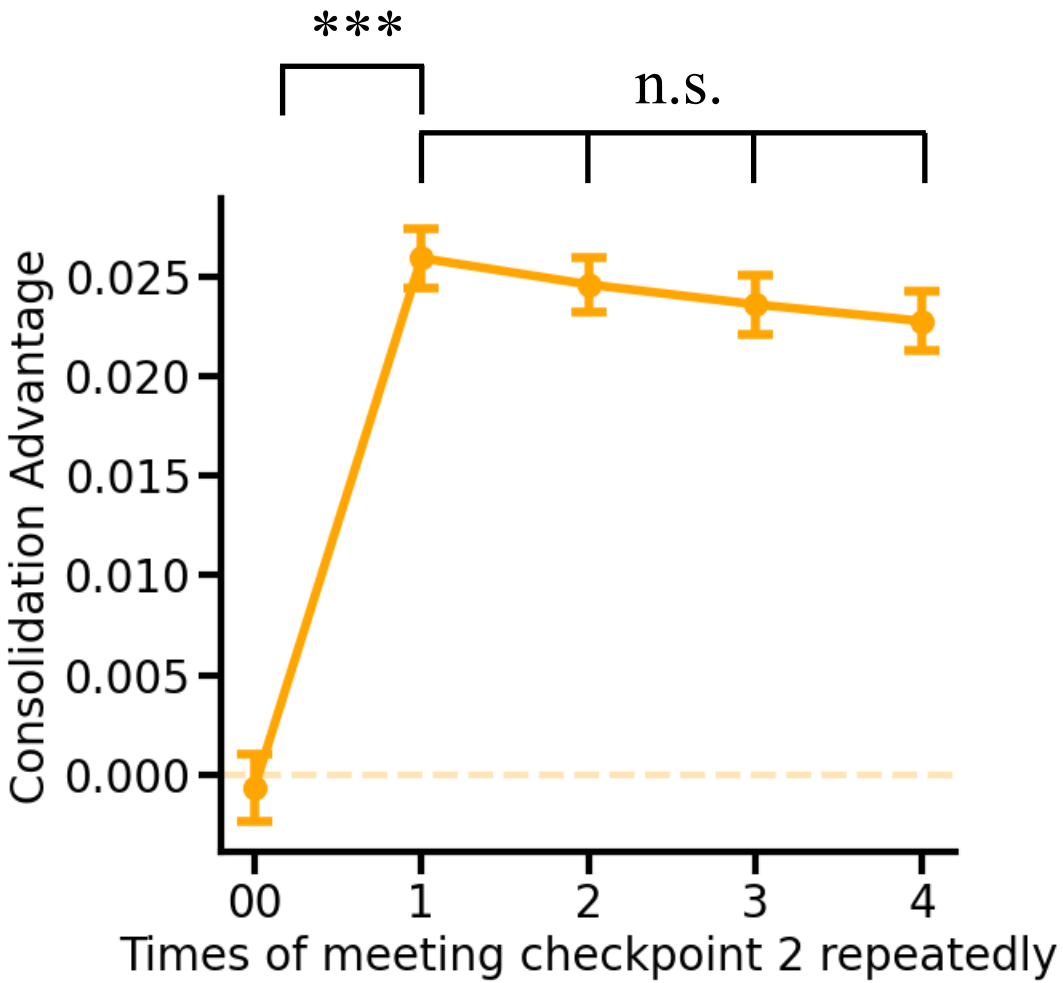
Value advantages of checkpoint 2 over checkpoint 1. Leftmost is the advantage before the checkpoint changes, and the others corresponds to different times the agent meets the same checkpoint repeatedly. (Significance test: ANOVA)

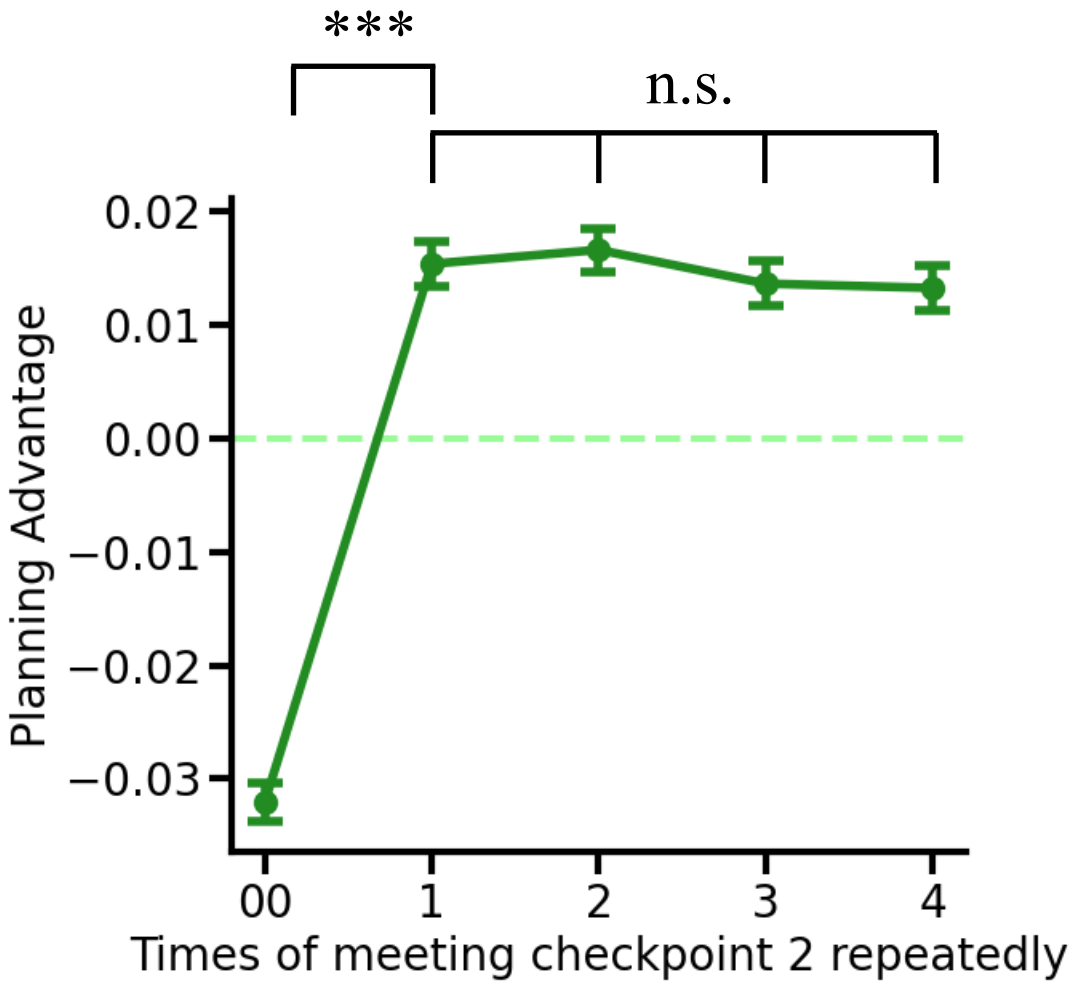
Experiments



Green line: delta of value advantages of checkpoint 2 over checkpoint 1.
Blue line: delta of advantages between all pairs of positions (25*25).
(Significance test: t-test)

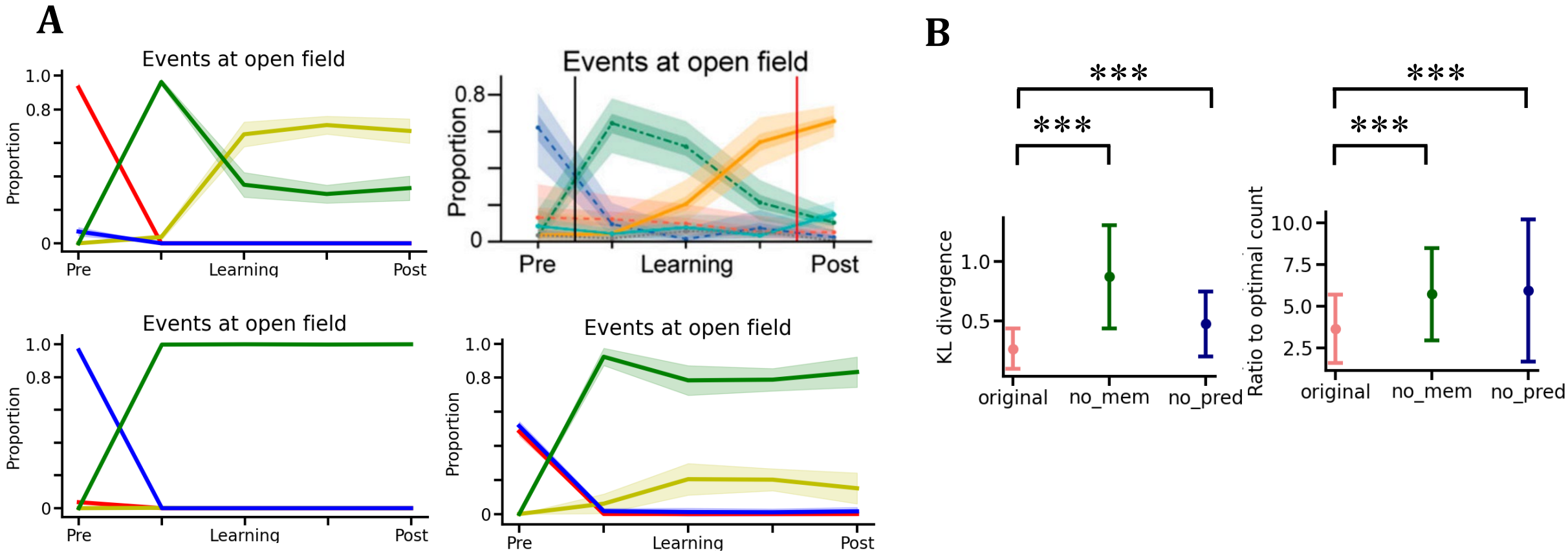
Experiments





Sum of values of different segments. The meaning of the color is the same as in the task diagram.

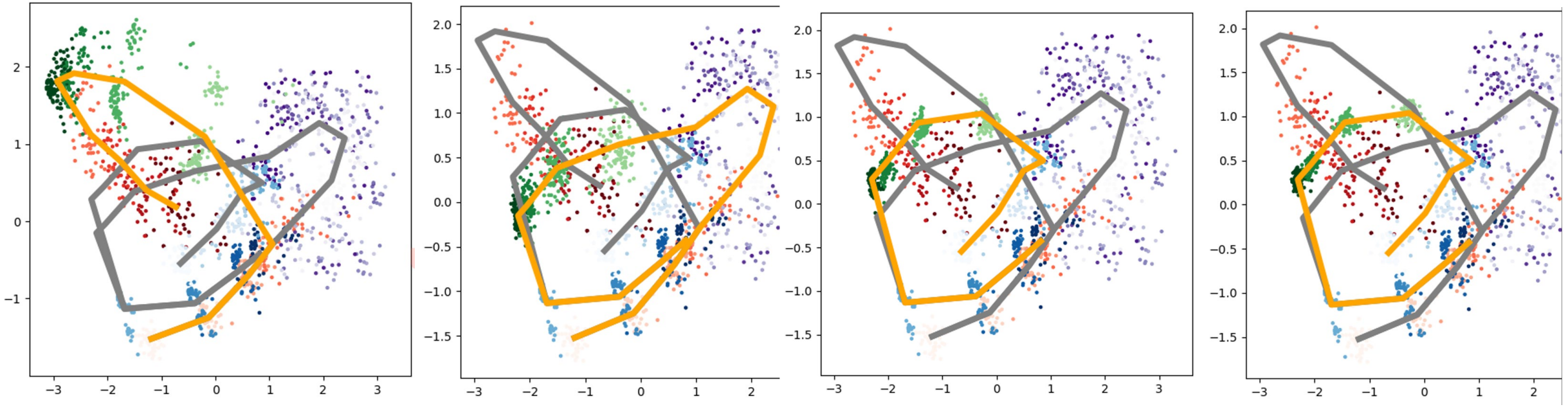
Experiments



Ablation Studies: After removal of either the memory objective function or path-integrating objective function, the replay curve is less similar with the original experimental curve (A: bottom left, remove memory, bottom right, remove path integration; B: distribution similarities).
The main difference is that the replay sequences representing shortcut to new checkpoint doesn't increase, which indicates the failure of hippocampus activity to capture the new checkpoint.

Experiments

initial 1 2 3

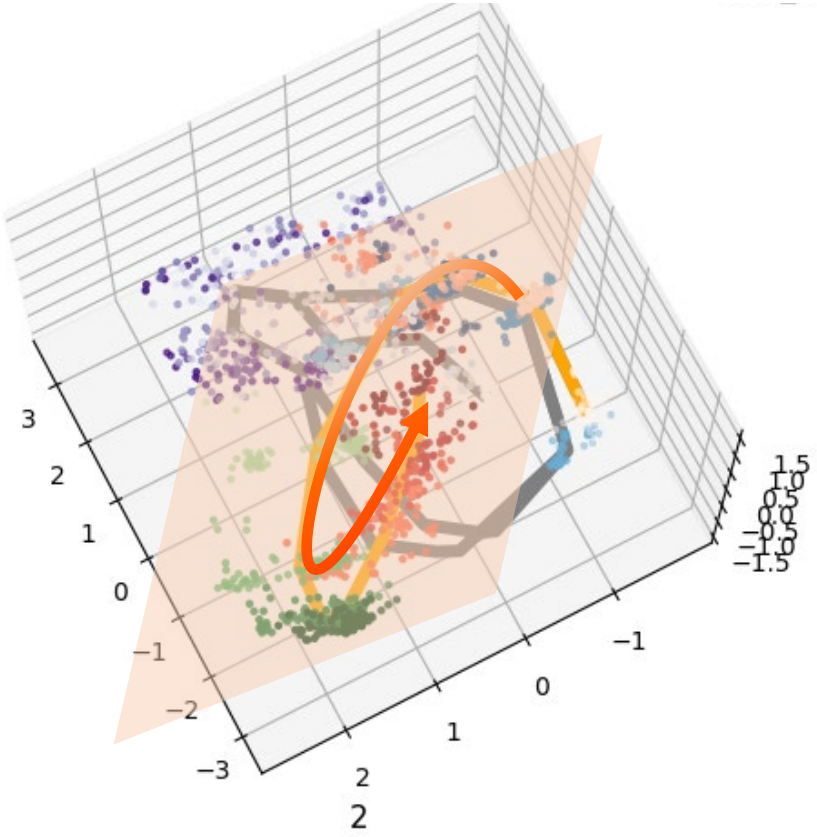


Manifolds: Principal components reduced from hidden states in PFC (2-dimension).
(initial: before replay, 1-3: different steps during replay).

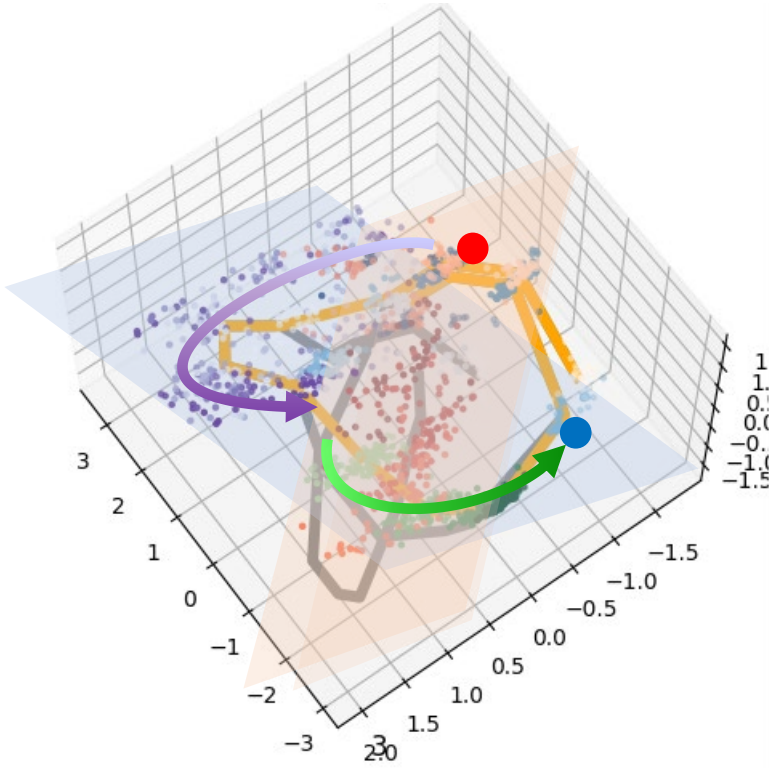
Experiments

Manifolds: Principal components reduced from hidden states in PFC (3-dimension).

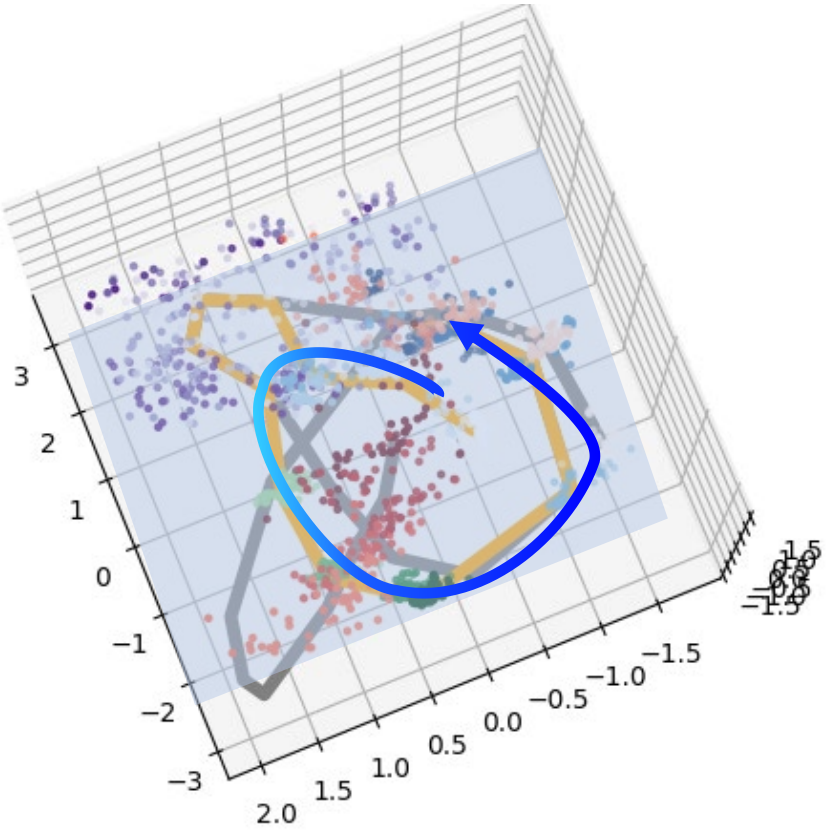
Reward at C1



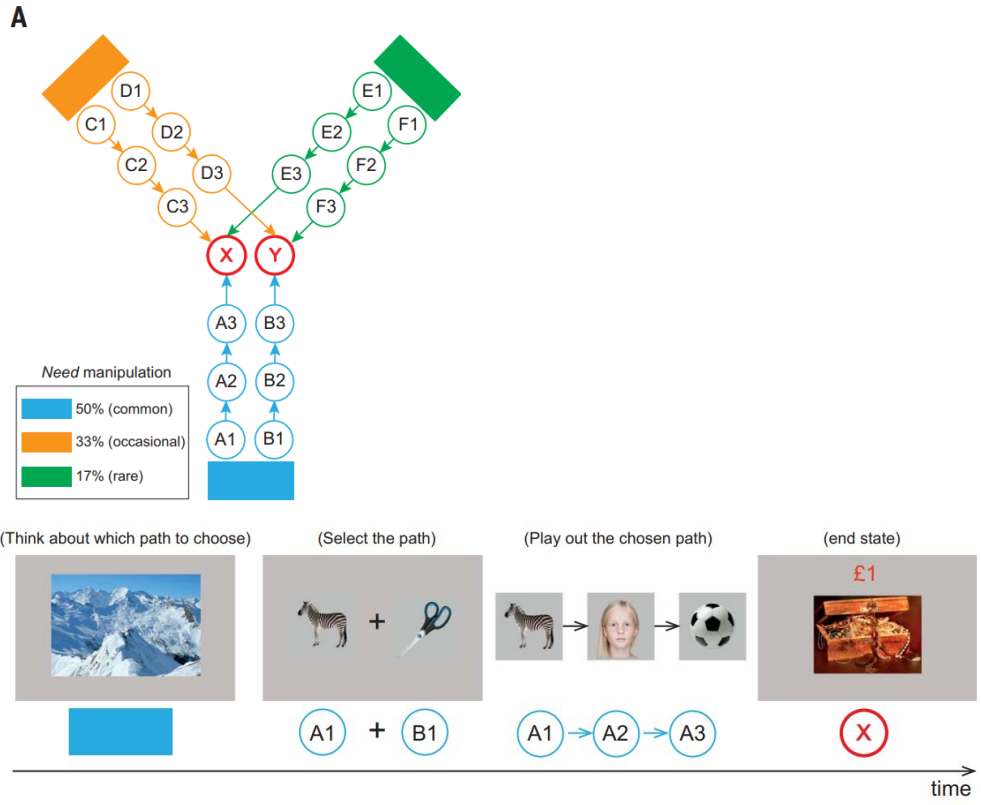
Transition



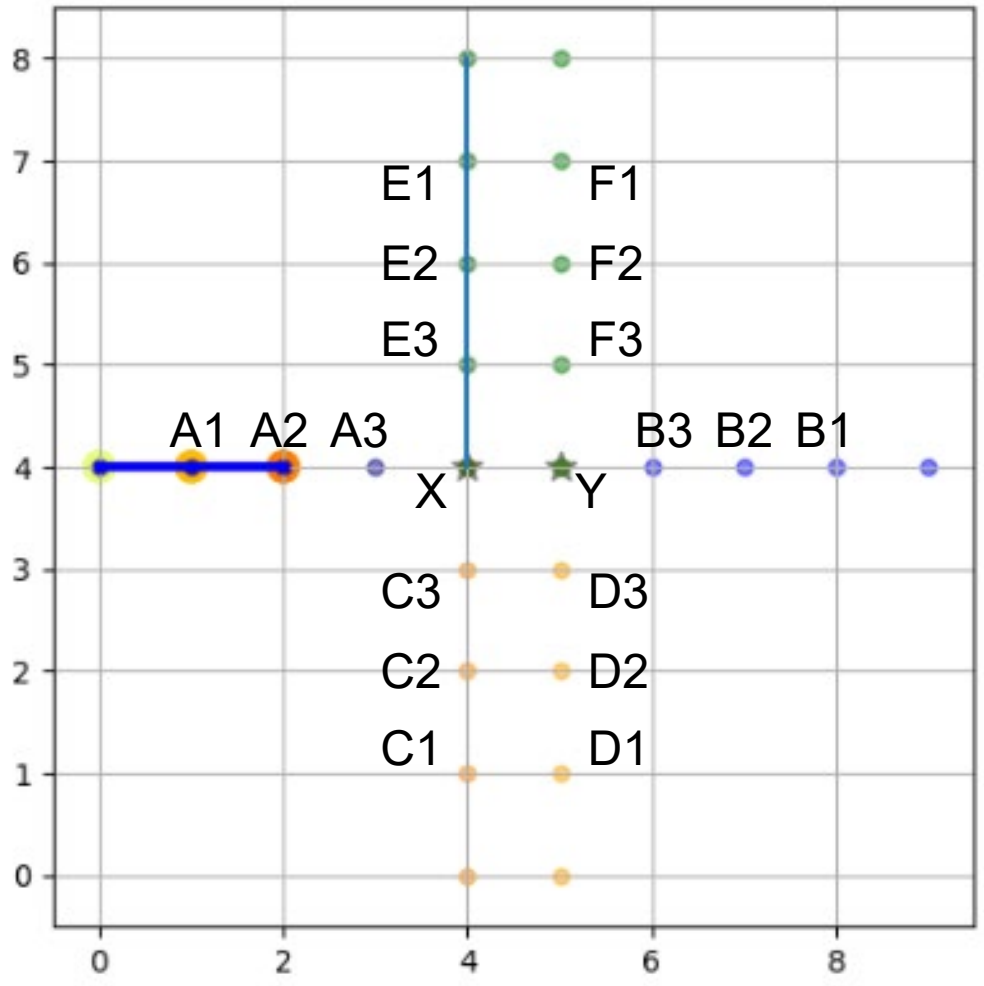
Reward at C2



Experiments (Humans)

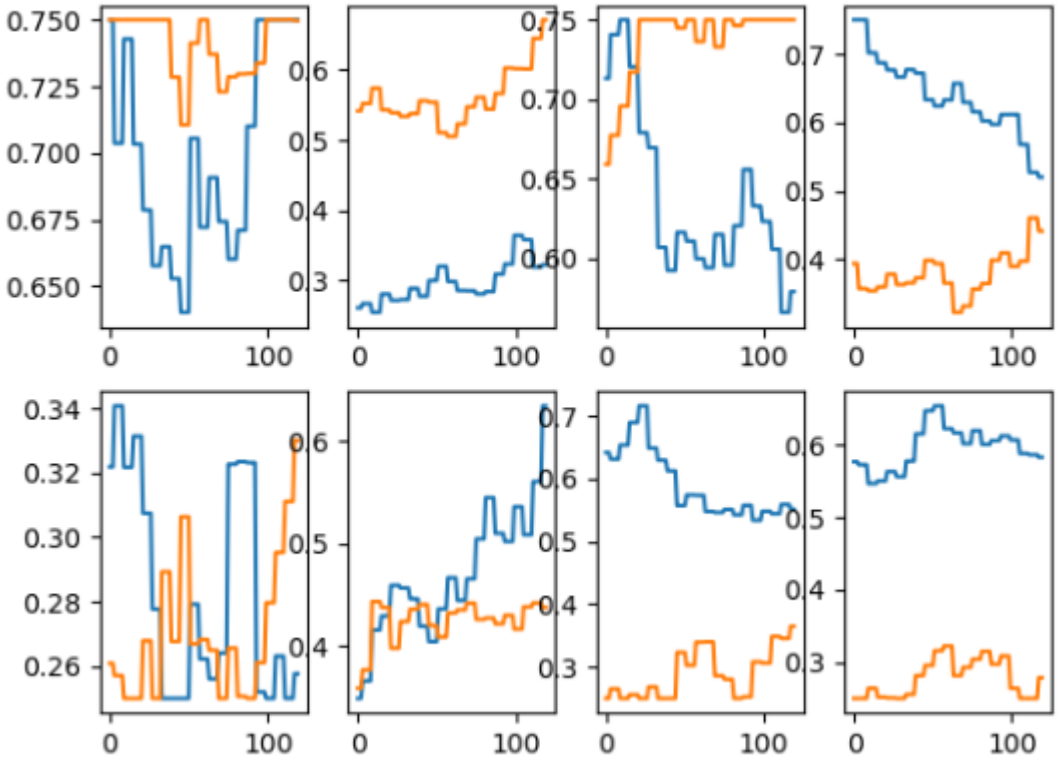


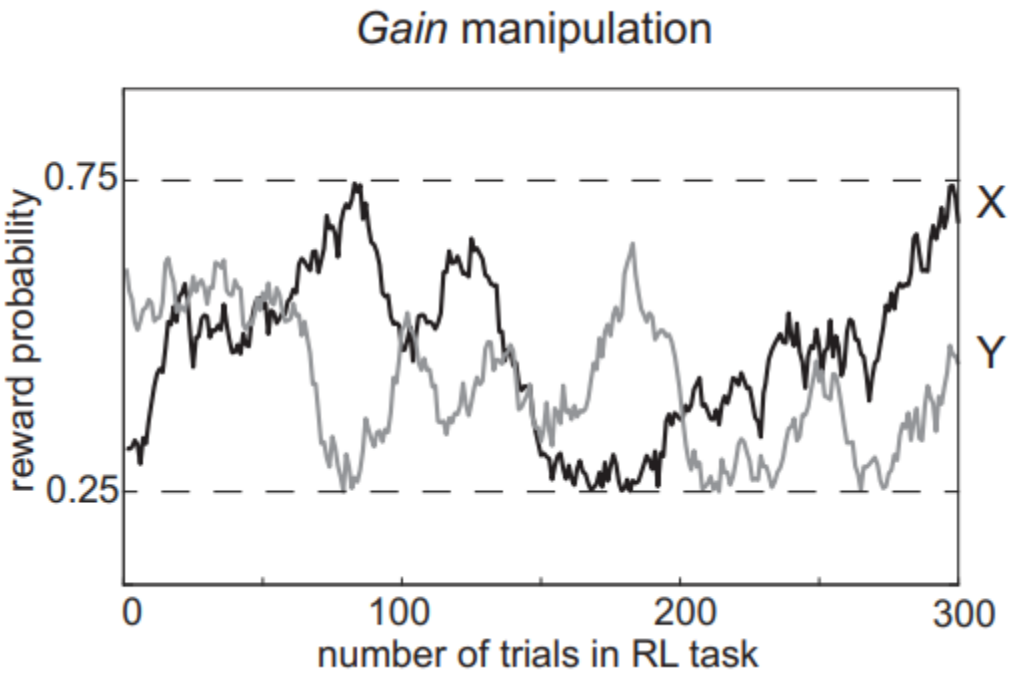
B



A: Setting of adaptive decision-making task. Each dot represents a path state in (Yunzhe Liu et al. 2021). The states with the same color constitute an arm, and every arm has two segments determined by the choice of the agent. B: the light blue line indicates the real "trajectory"; the dark blue line indicates the replay sequences. Here, the agent is carrying out non-local replay.

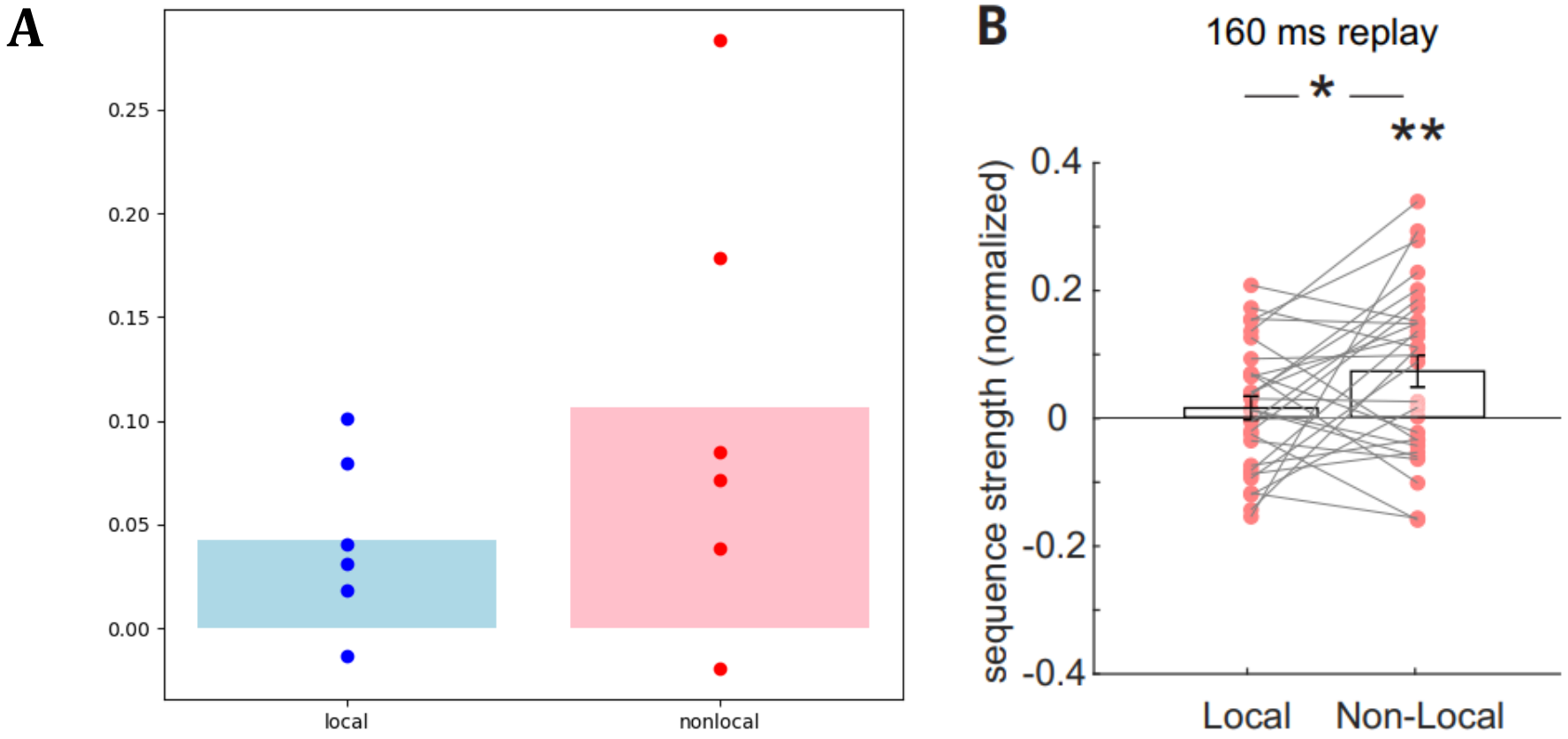
Experiments





Gain modulation in the task. The same setting (Gaussian random walk with standard deviation=0.025) as in the original article is used. Left: ours (orange: x; blue: y); Right: original paper.

Experiments



Sequence strength (a metric from the original paper) of local and non-local backward replay in our model resembles that of 160-ms backward replay in human. This shows that the backward replay in our model also encodes nonlocal experiences as opposed to local experiences. Left: ours; Right: original paper.

Limitations and Future works

(1) Use “plane” (ANN) to study “bird” (brain). Biologically interpretable models are more convincing especially for biologists.

(2) Time compression and dilation is important for replay (time scale of replay), but our RNN based model is sensitive to these.

(3) Internal model means the on-going activities but not the physical plasticity (short-term plasticity, replay induced plasticity), thereby sensitive to interrupting (but humans not).

(4) Hippocampus as a cognitive map may provide insights for developing navigation functions in LLM.

Thank you!