Shifts in learning strategies underlie rodent behavior during dynamic foraging

Nhat Minh Le[1], Murat Yildirim[2], Hiroki Sugihara[2], Yizhi Wang[1], Mriganka Sur[1]
[1]Department of Brain and Cognitive Sciences, [2]Picower Institute for Learning and Memory, MIT, Cambridge, MA

In uncertain foraging environments, a critical problem agents face is how to adapt their learning in accordance with the uncertainty of the environment and knowledge of the hidden structure of the world. In this context, previous studies distinguished between two modes of learning: model-free learning updates action values while balancing between exploration and exploitation, while inference-based behavior leverages knowledge of world's dynamics to infer the current state of the world. However, it is unclear how these behavioral strategies can be cleanly distinguished based on behavioral data and how animals transition between these modes during training. Here, we tackled these questions by examining rodent behavior in a dynamic foraging task. We trained head-fixed mice to select between two choices, left and right wheel turns. Only one direction was rewarded in each trial, and the target direction changed in blocks of 15-25 trials. To determine whether rodent behavior was better described by the model-free or inference-based strategies during learning, we simulated choice sequences of model-free and inference-based agents to gain insights into their behavioral signatures. Our model-free agent implements a Q-learning algorithm, while the inference-based agent performs Bayesian inference assuming a 2-state model of the world. We identified four criteria to characterize the behavior of these agents and used them to fit rodent behavioral data: the slope and offset of their switching decisions, the degree of exploration, and the relationship between switching and the performance in the previous block. Across all animals, we observed a consistent decrease in the times to switch the wheel turn direction, and a decrease in the degree of exploration with training. By performing a model comparison using log-likelihoods, these observations were better explained by a transition from model-free to inference-based mode of learning, as opposed to an increase in learning rate in a model-free agent. A 3-state Hidden-Markov model was fitted to the rodent behavioral data, revealing transitions between left-exploit, right-exploit and exploratory states. The duration of exploratory states decreased with training, consistent with the transition to the inference-based strategy. These analyses and characterizations form the basis of understanding shifts in behavioral strategies during dynamic foraging. Together with widefield calcium imaging across the whole cortex during the task, they will be important for identifying unique neural signatures of global brain states that are associated with distinct modes of learning.

Key words: Reward, Exploration, Learning

Theme/Topic: H.08.f. Learning and Memory - Striatal and corticostriatal circuits, H.03.a. Decision Making - Computational models

Support: NIH grants 1R01MH126351, EY028219, K99EB027706, Paul and Lilah Newton Brain Science Research Award