

How does the brain learn from feedback?

Readings for today

- Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual review of neuroscience*, 35, 287-308.

Topics

- Actors and critics in the brain
- Representation of value

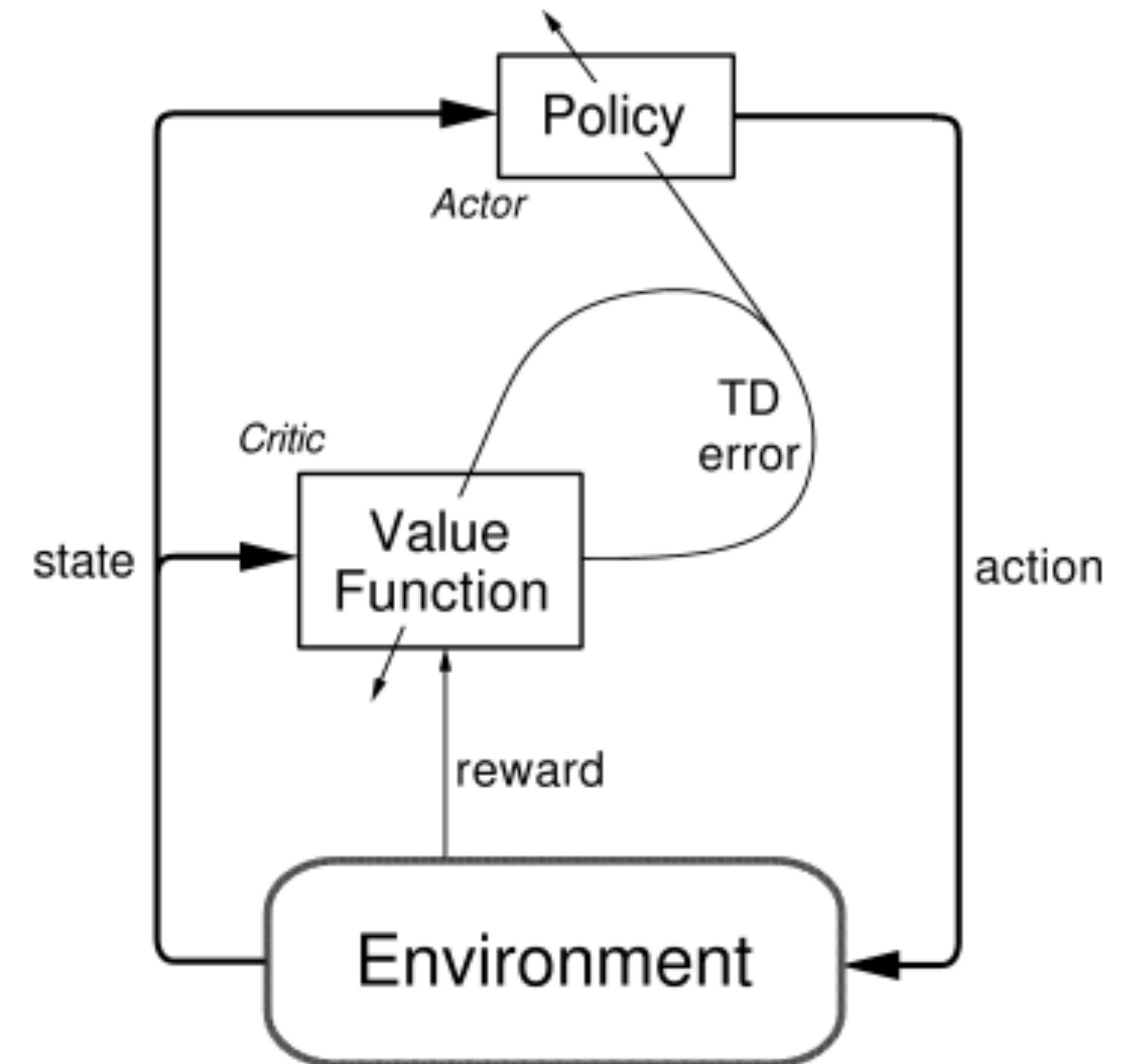
Actors and critics in the brain

Actor-Critic Learning

Two interacting processes (networks):

- **The actor:** decides which action should be taken.
- **The critic:** informs the actor how good the action was and to adjust adjust.

The learning of the actor is based on policy gradient approach (i.e., RL)

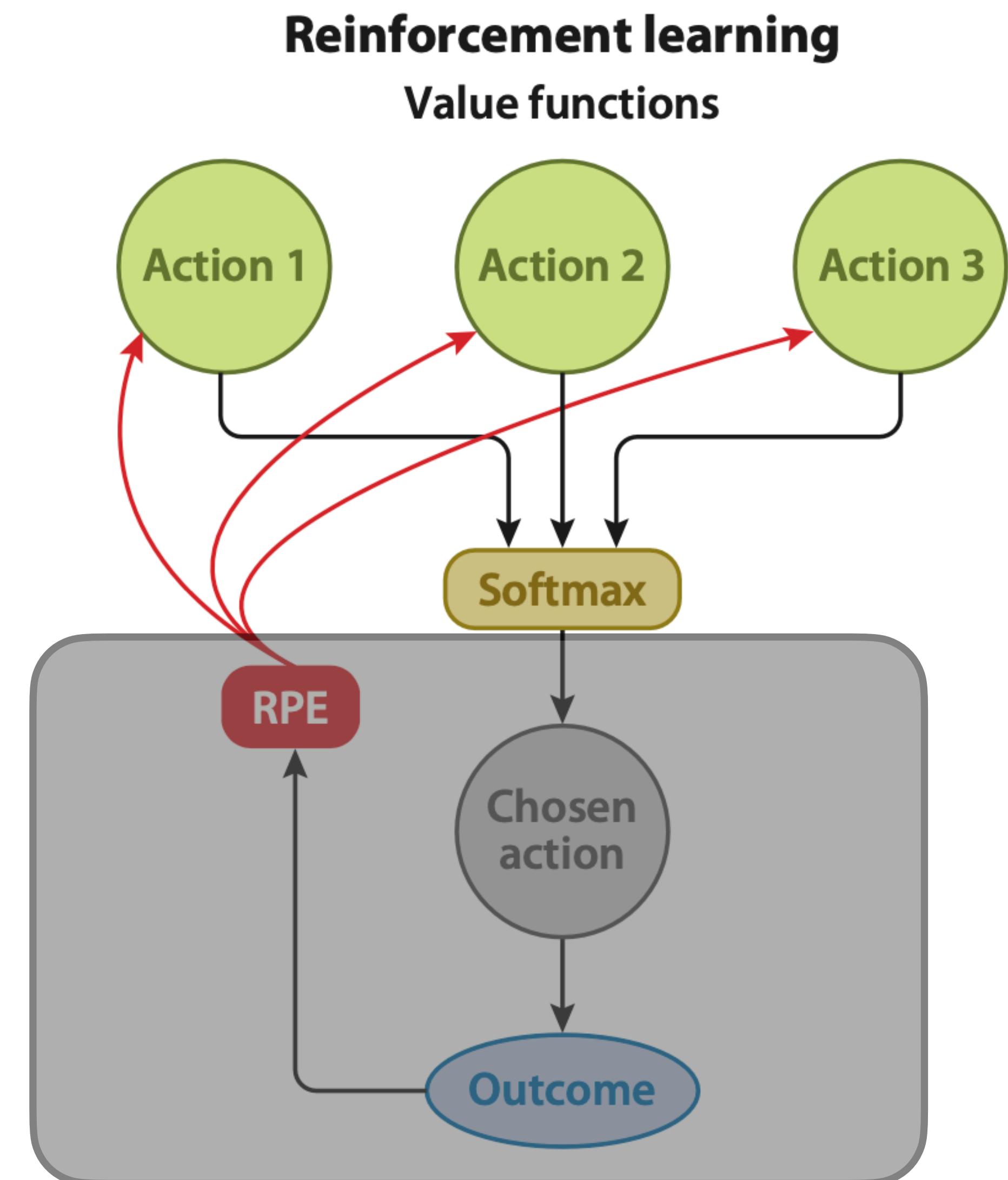


The reward prediction error

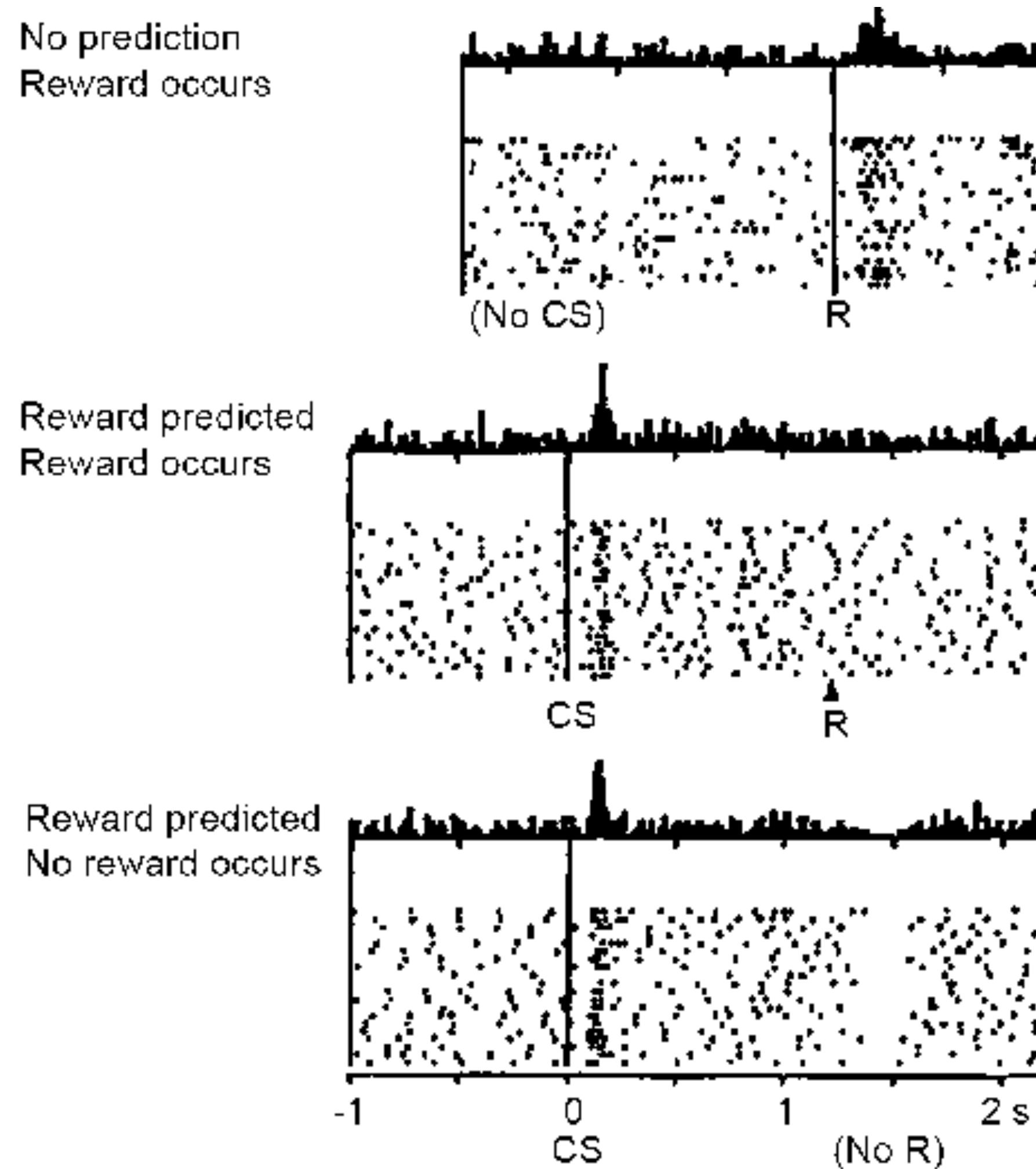
Reward prediction error (RPE)

$$Q_{n+1} \doteq Q_n + \alpha [R_n - Q_n]$$

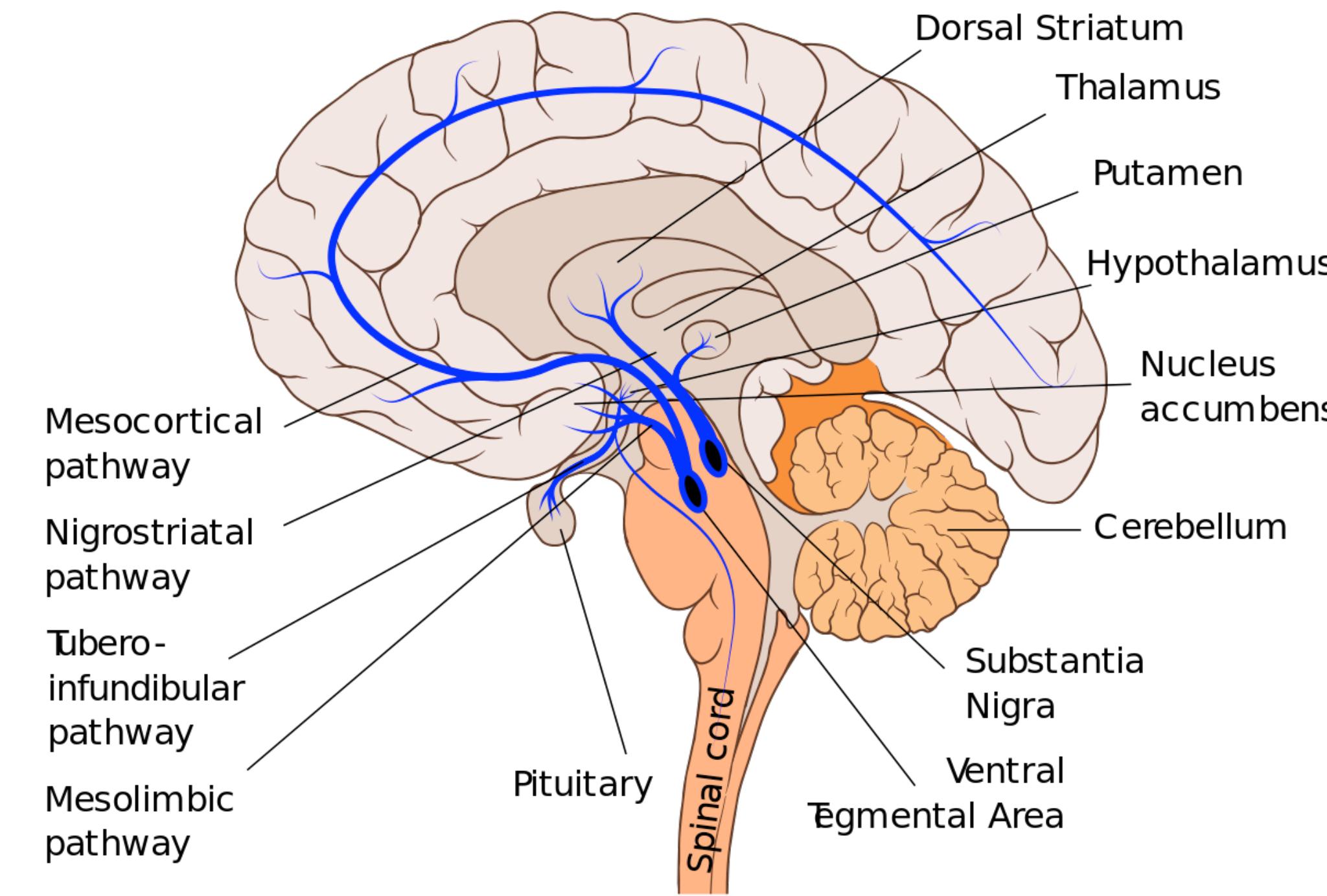
A concept that measures the difference between expected and received rewards, signaling the need to adjust behaviors or expectations when outcomes do not match predictions.



Phasic dopamine tracks with RPEs

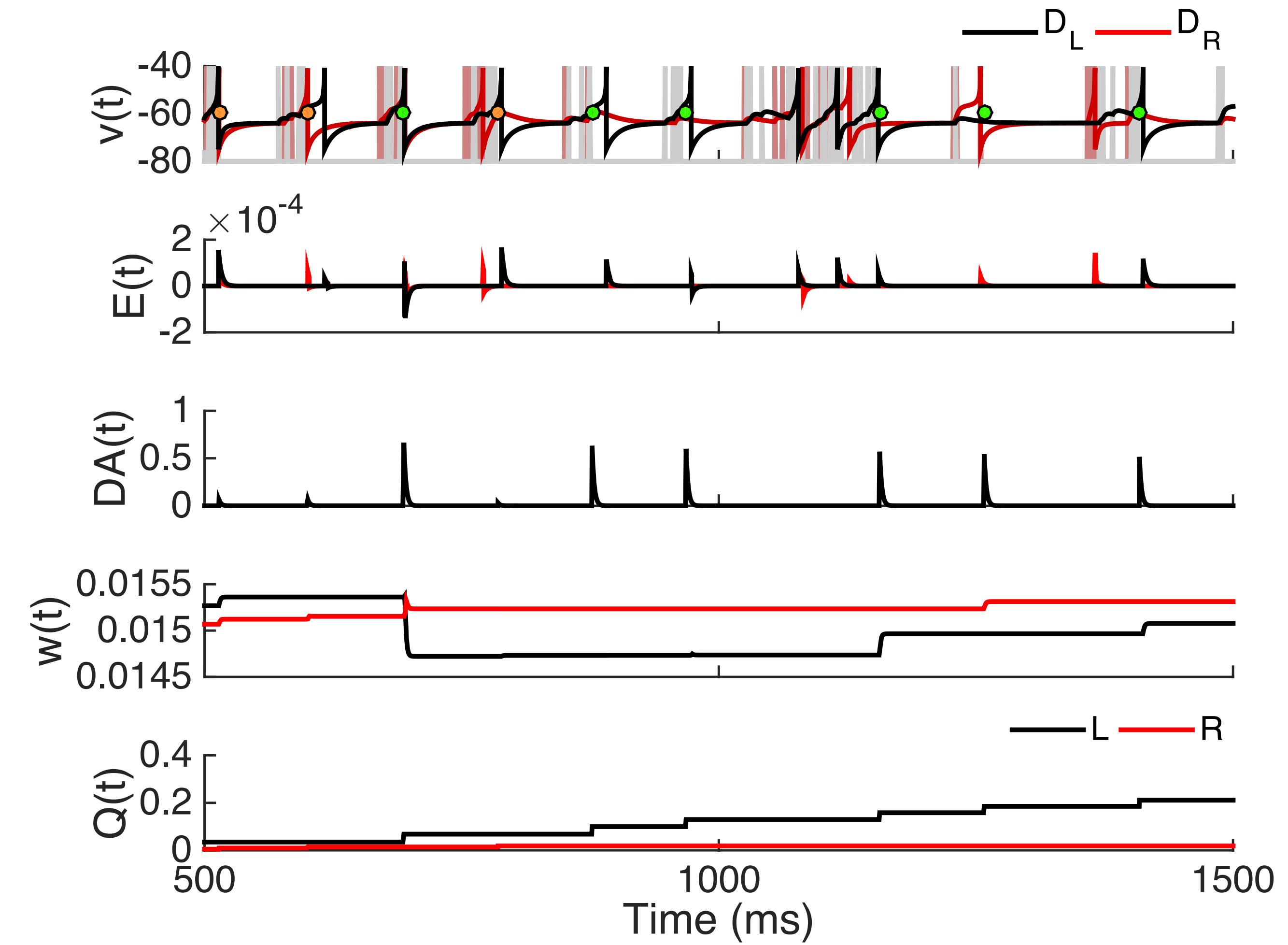
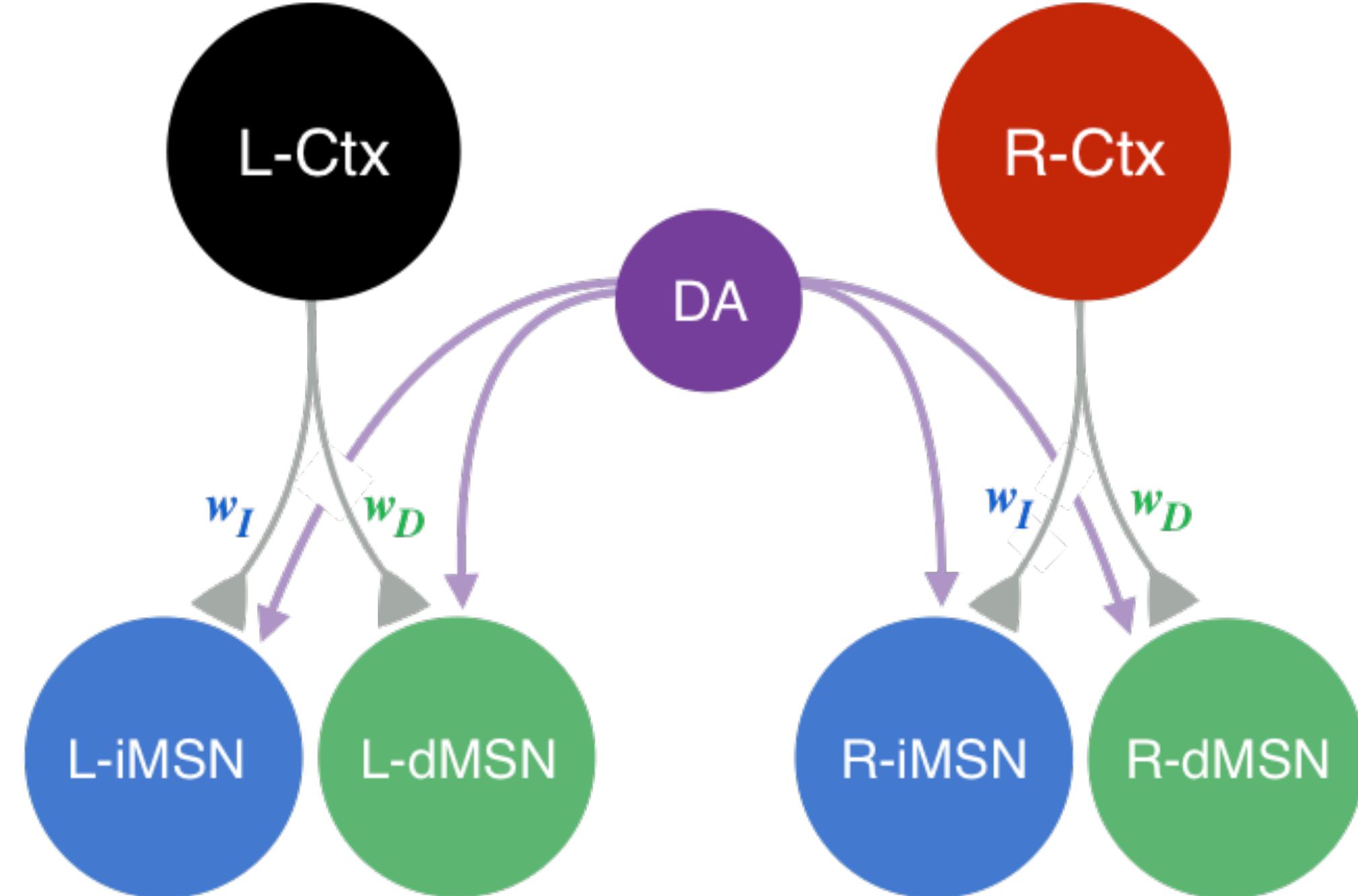


Cells in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) both show activity tied to the reward prediction error.

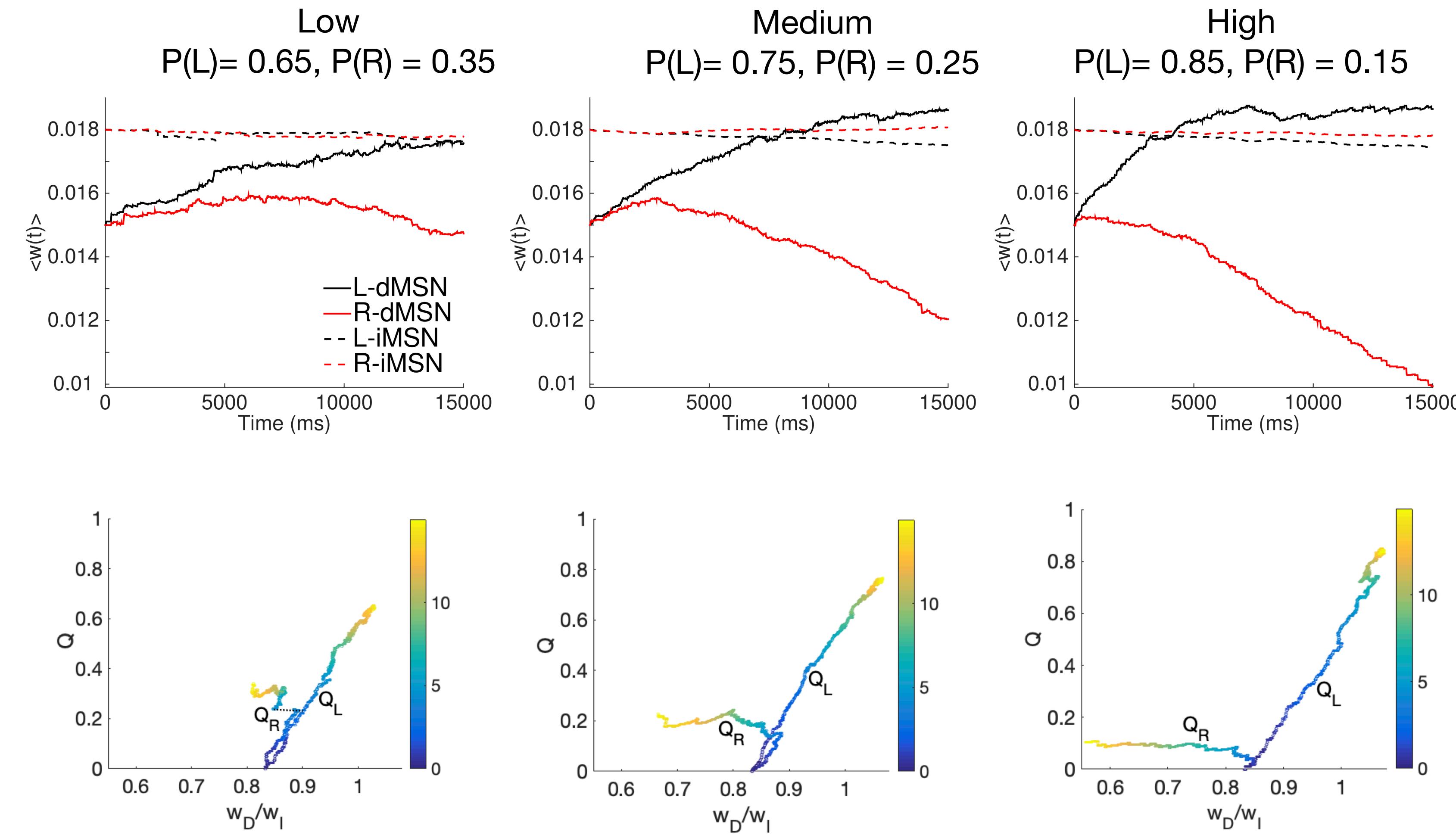


https://en.wikipedia.org/wiki/Dopaminergic_pathways

Spike-timing dependent plasticity



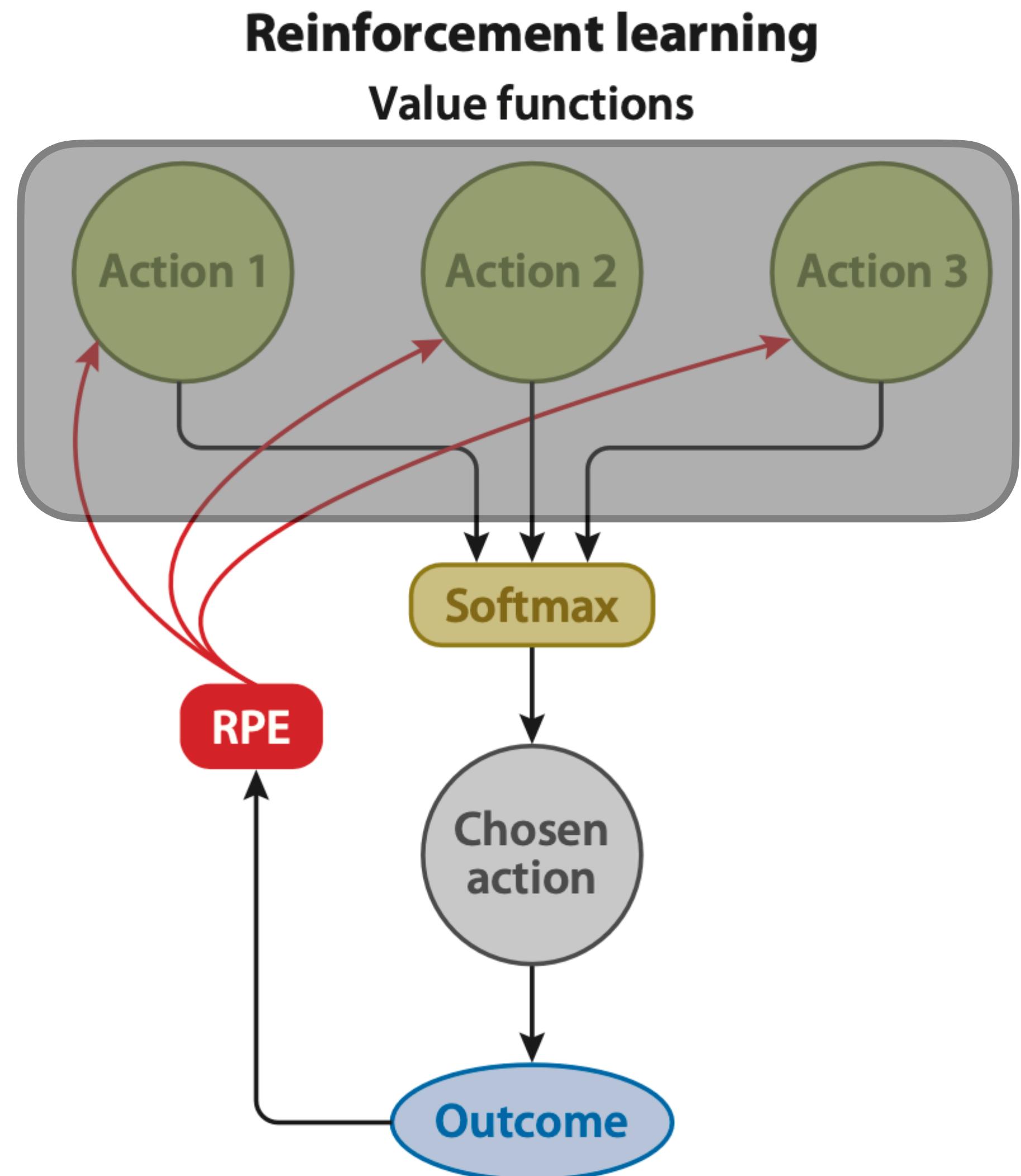
Temporal difference update



Phasic dopamine signals at the corticostriatal synapses can update behavior consistent with tracking action value (Q)

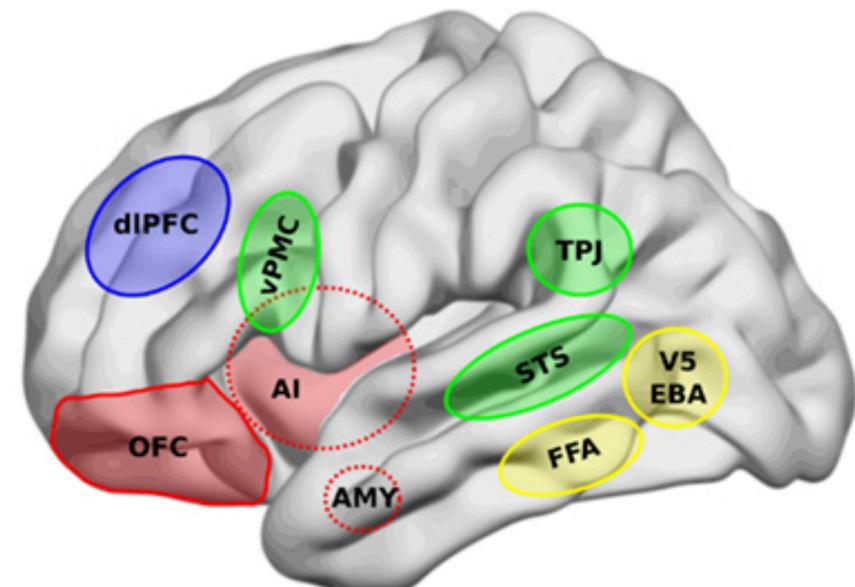
Representation of value

Where is value represented?

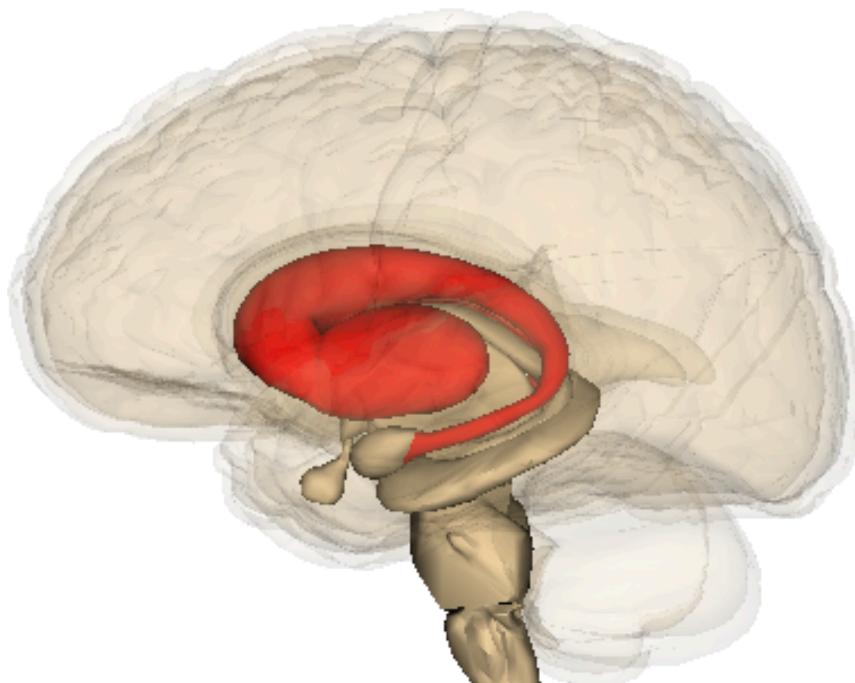
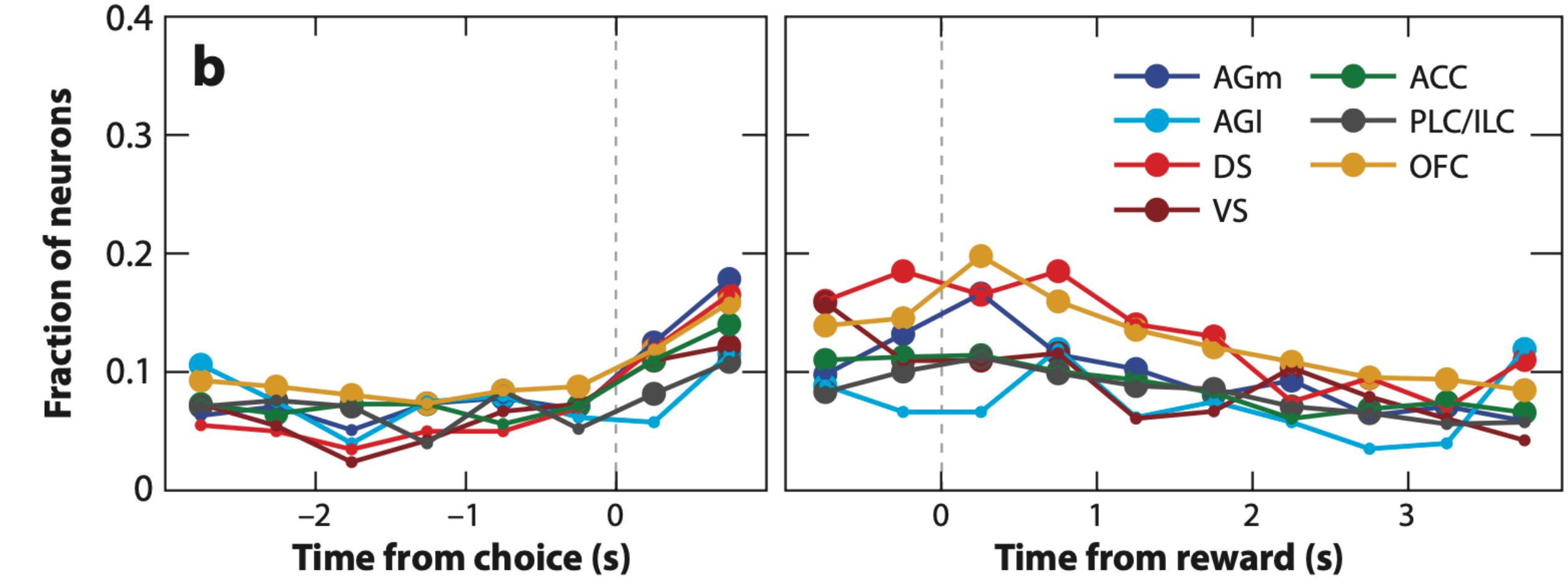
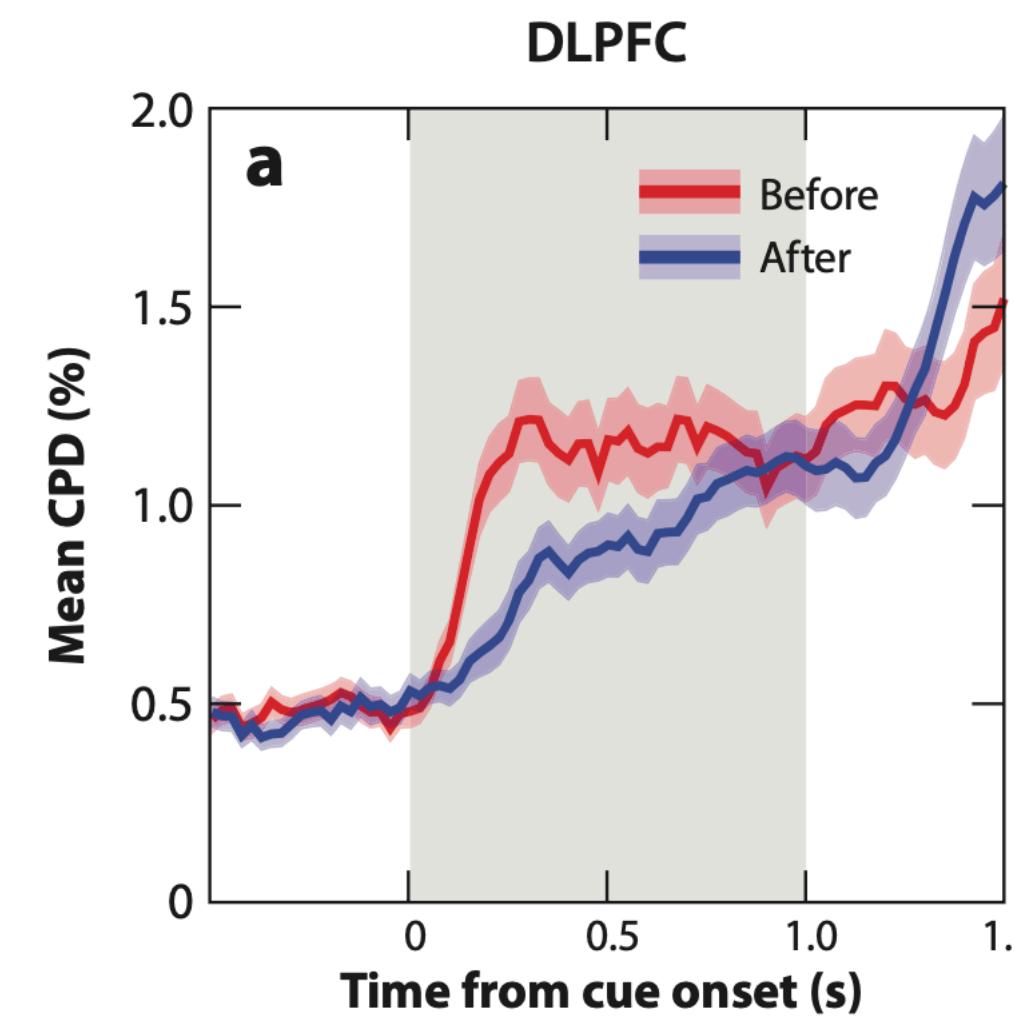


If RPE's drive an adjustment in the value of possible actions, then where does the brain represent (and update) action values?

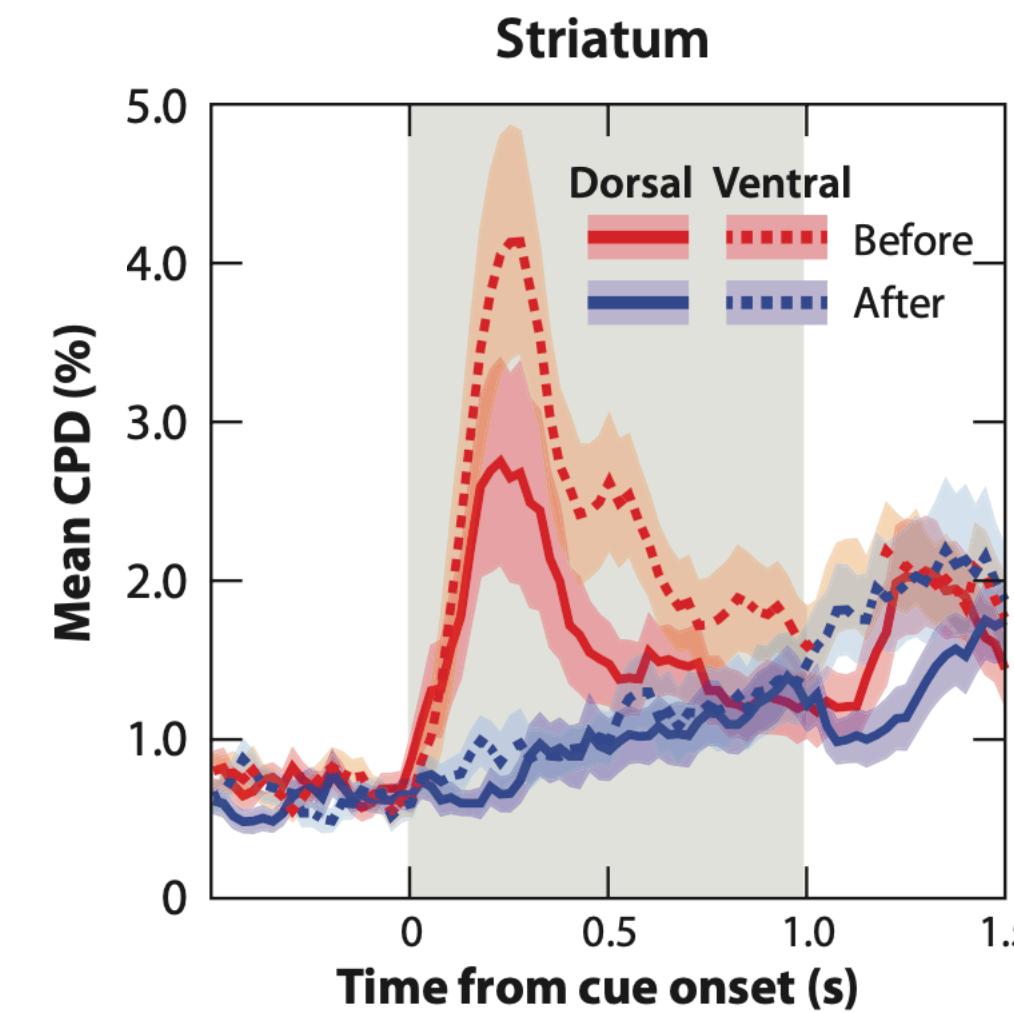
Neural responses to value after training



(Credit: Wikipedia)

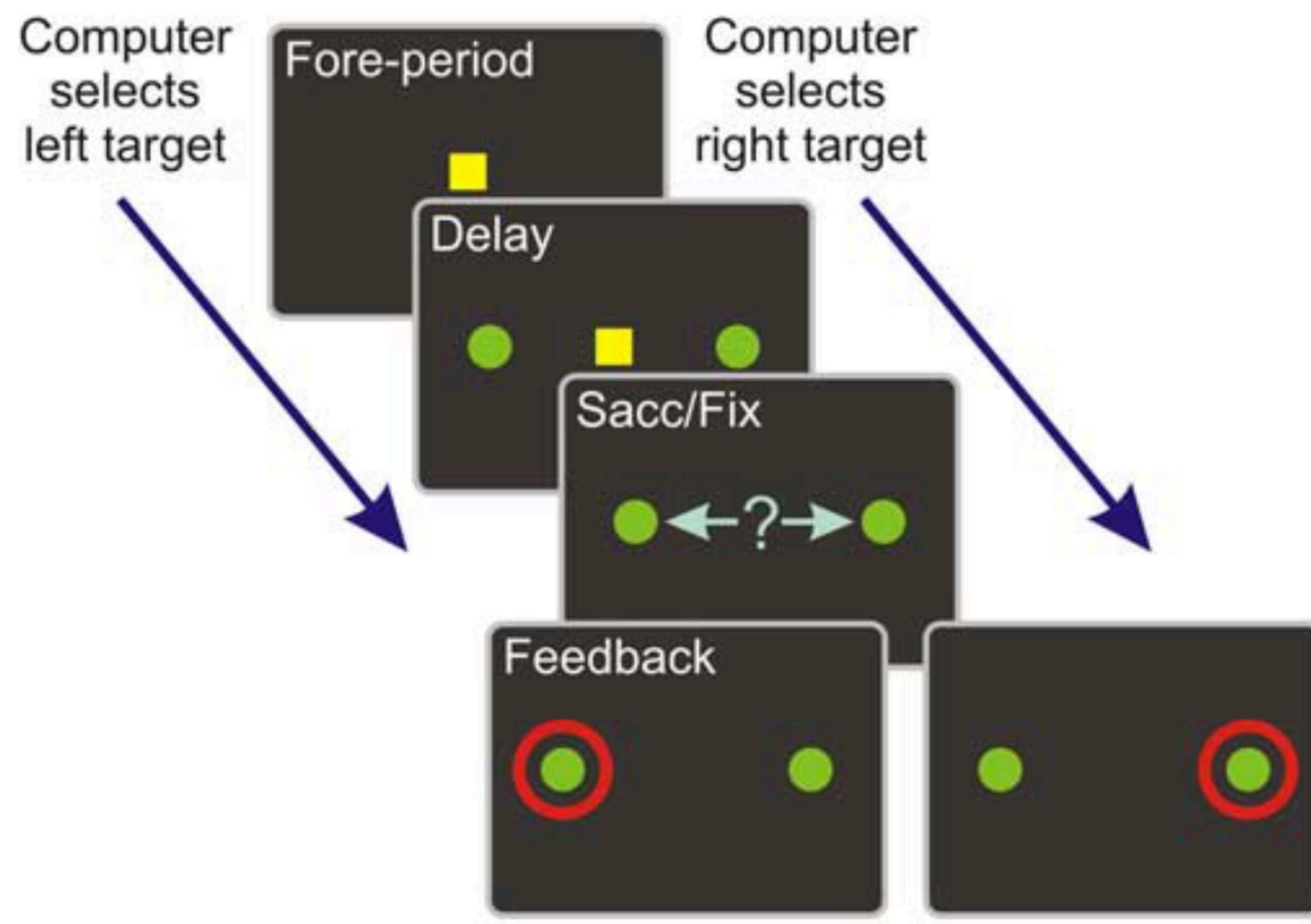


(Credit: Wikipedia)



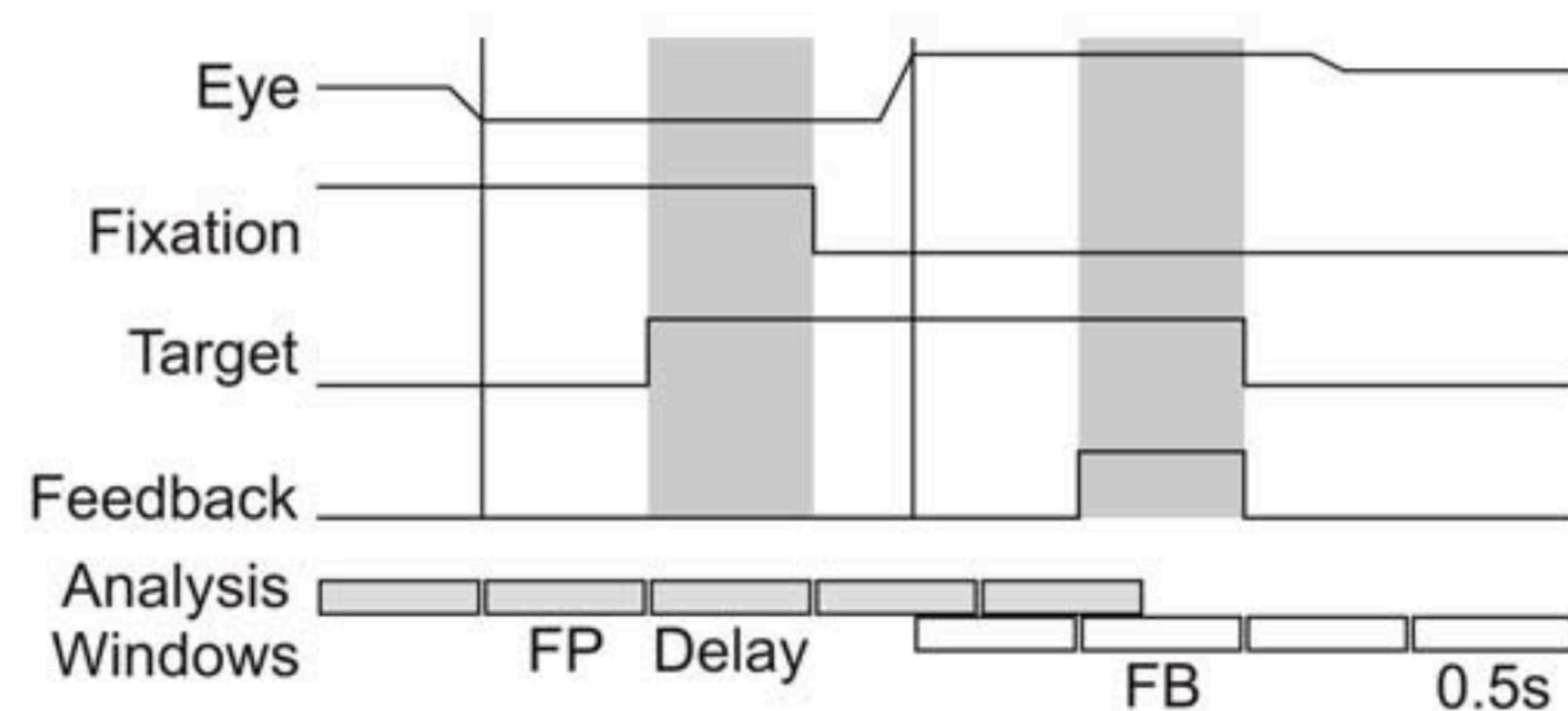
Training in a simple binary choice task changes the neural responses to action cues, in both DLPFC (down) and striatum (up; Ventral Striatum). Between 10-20% of neurons in various cortical and striatal regions are tuned to the Q-value for an action.

The matching pennies task



The matching pennies task is a variant of the bandit task where a computer tracks the players performance to estimate their action values to each target and either give feedback consistent with their value (exploitation) or against the value (exploration).

The task has two distinct periods useful for probing underlying neural representations.



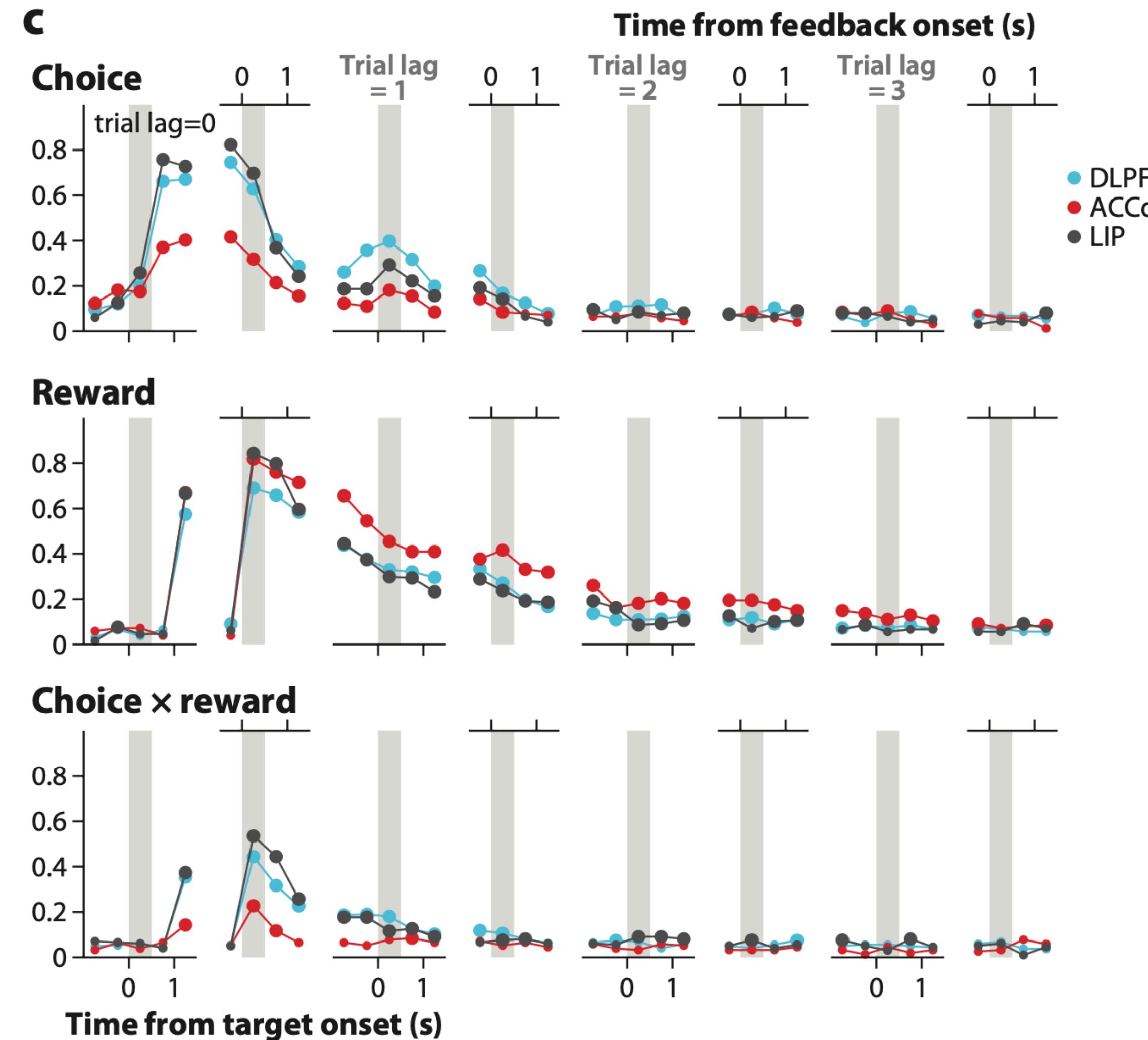
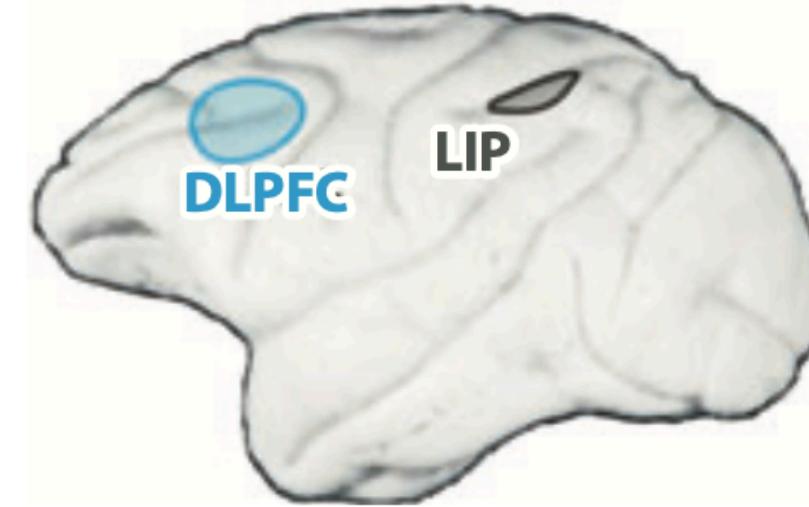
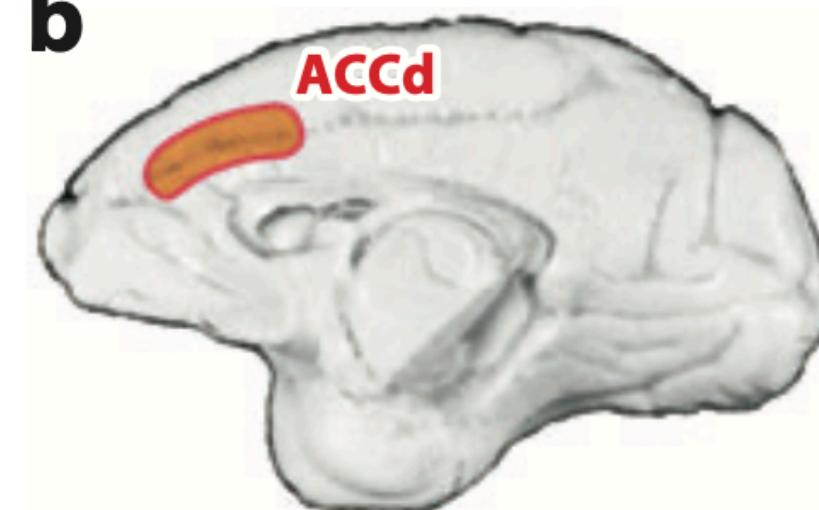
- **Choice period**: when an action is selected
- **Feedback period**: when a reward is delivered

Representing choice, value, or both?

a Matching-pennies task

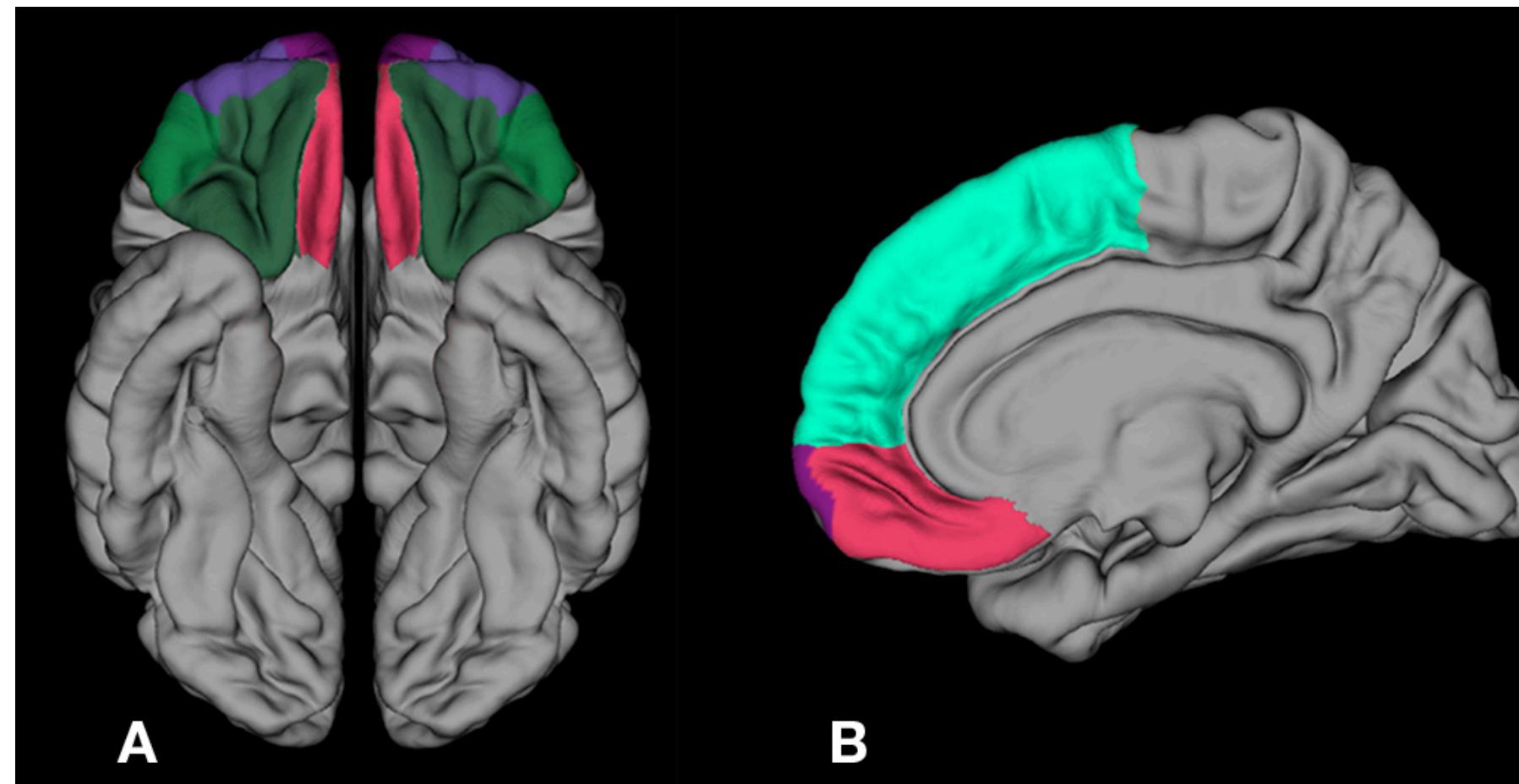


b



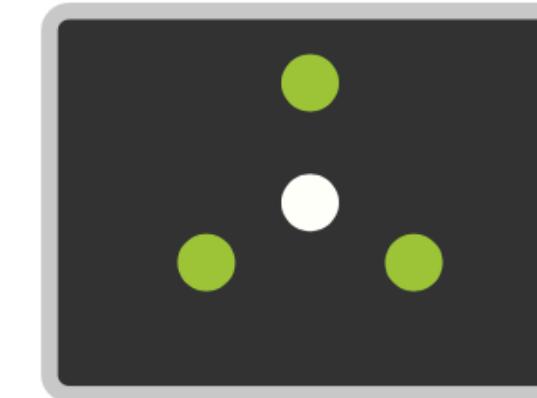
- Lateral prefrontal and parietal regions had a higher proportion of cells tuned to choice.
- Medial prefrontal regions (ACC) had a higher proportion of cells tuned to reward in later trials.

Orbitofrontal cortex reflects RPEs and value

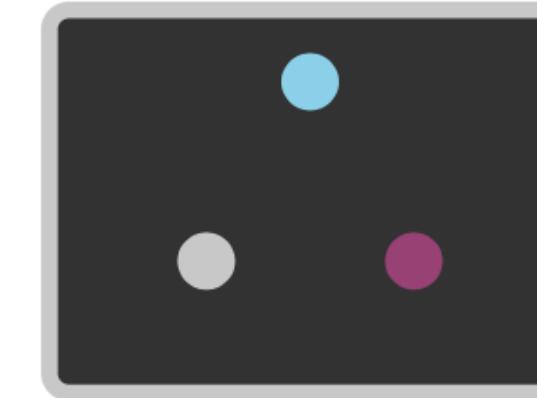


Rock-paper-scissors task

Target onset

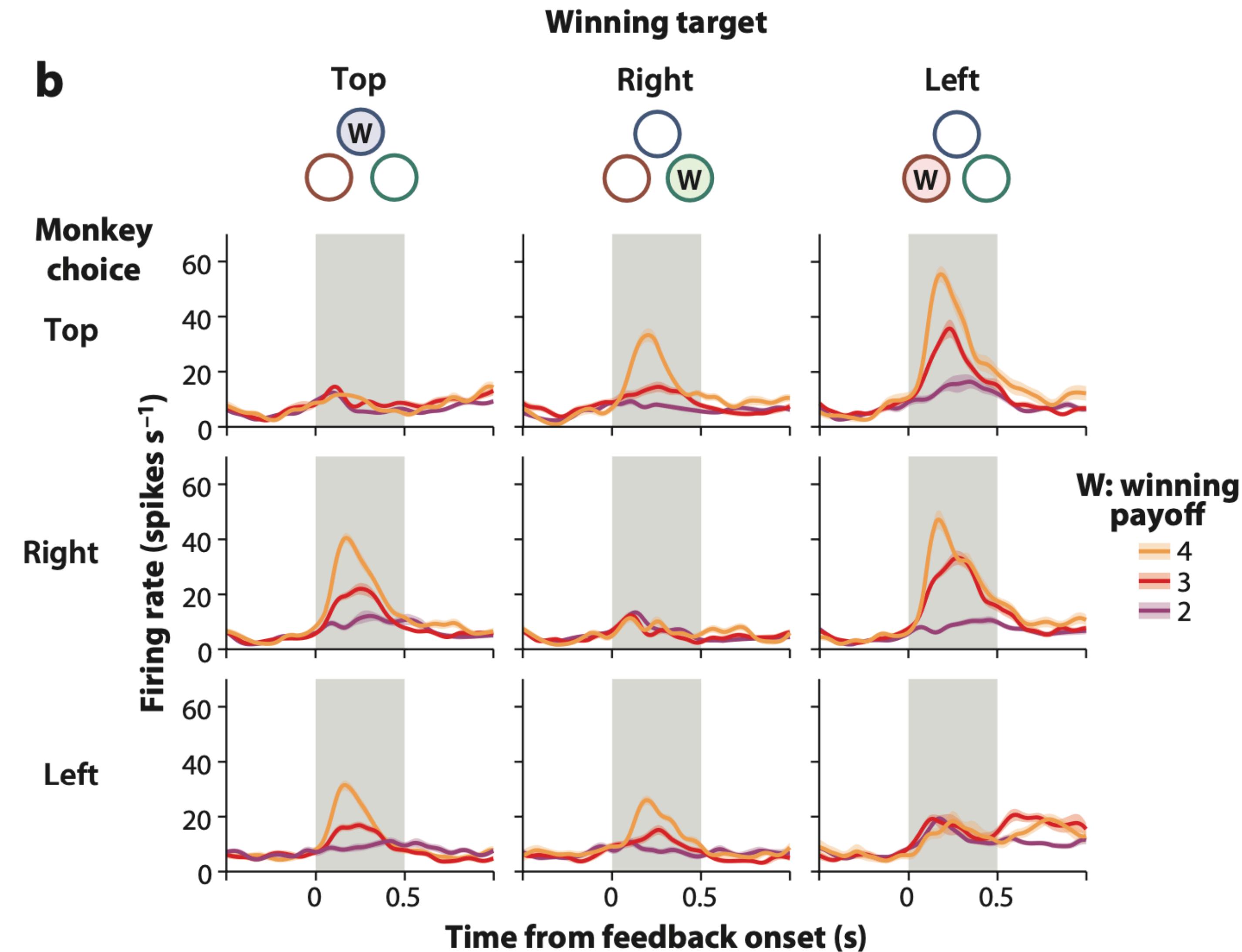


Feedback onset



Reward

- ✕
- ⚡
- 💧
- 🍃
- 🍃

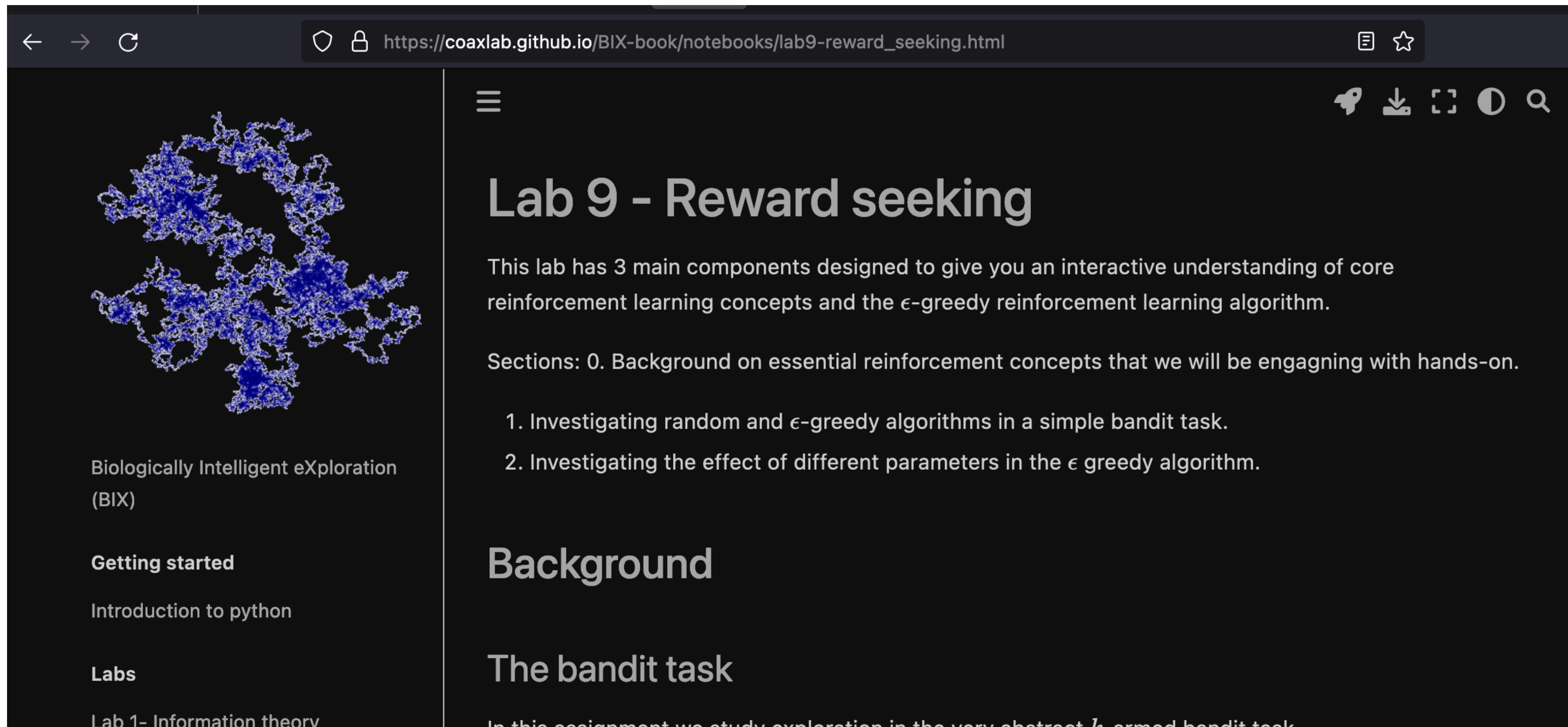


Take home message

- Phasic dopamine signals in the substantia nigra and ventral tegmentum track with the reward prediction error.
- Prefrontal cortex regions, along the striatum and some parietal regions, appear to keep track of action values and update their responses with feedback.

Lab 9: Reward seeking

URL: https://coaxlab.github.io/BIX-book/notebooks/lab9-reward_seeking.html



The page is titled "Lab 9 - Reward seeking". It describes the lab as having 3 main components designed to give an interactive understanding of core reinforcement learning concepts and the ϵ -greedy reinforcement learning algorithm. The sections listed are: 0. Background on essential reinforcement concepts that we will be engaging with hands-on; 1. Investigating random and ϵ -greedy algorithms in a simple bandit task; and 2. Investigating the effect of different parameters in the ϵ greedy algorithm.

Background

The bandit task

In this assignment we study exploration in the very abstract k -armed bandit task.

Navigation links on the left include: Biologically Intelligent eXploration (BIX), Getting started, Introduction to python, Labs, and Lab 1- Information theory.