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

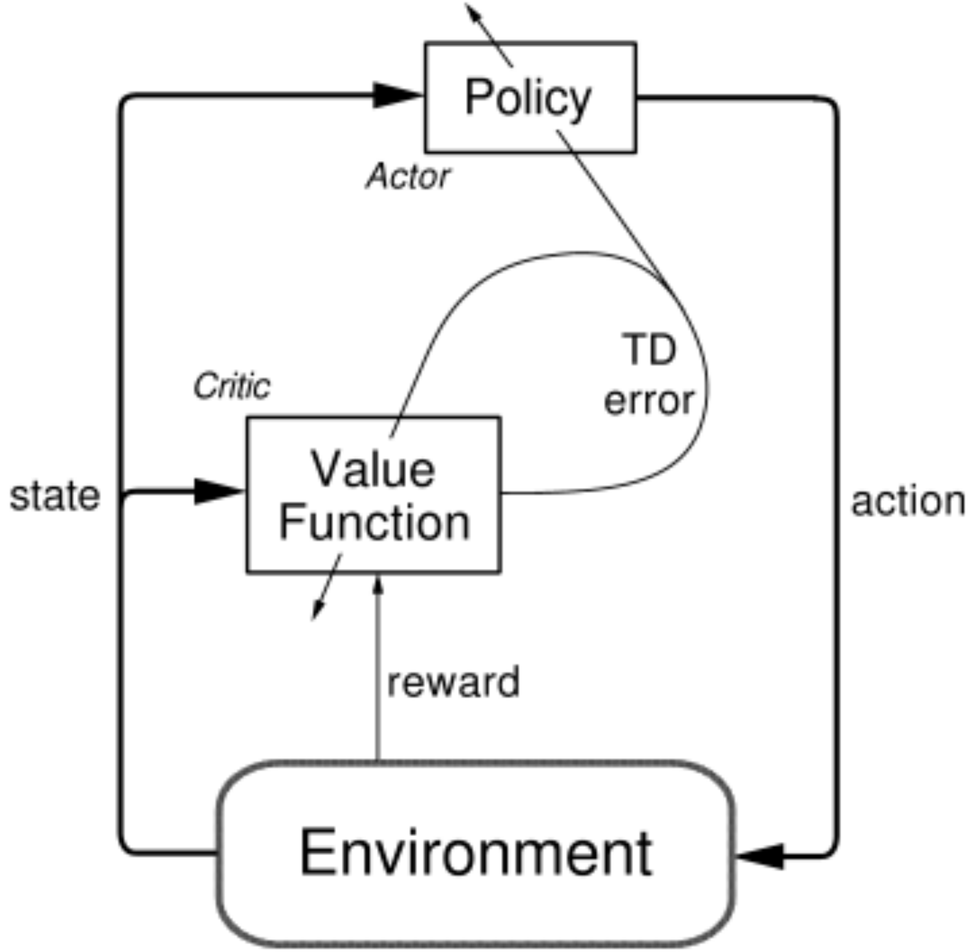
Actors and critics in the brain

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)

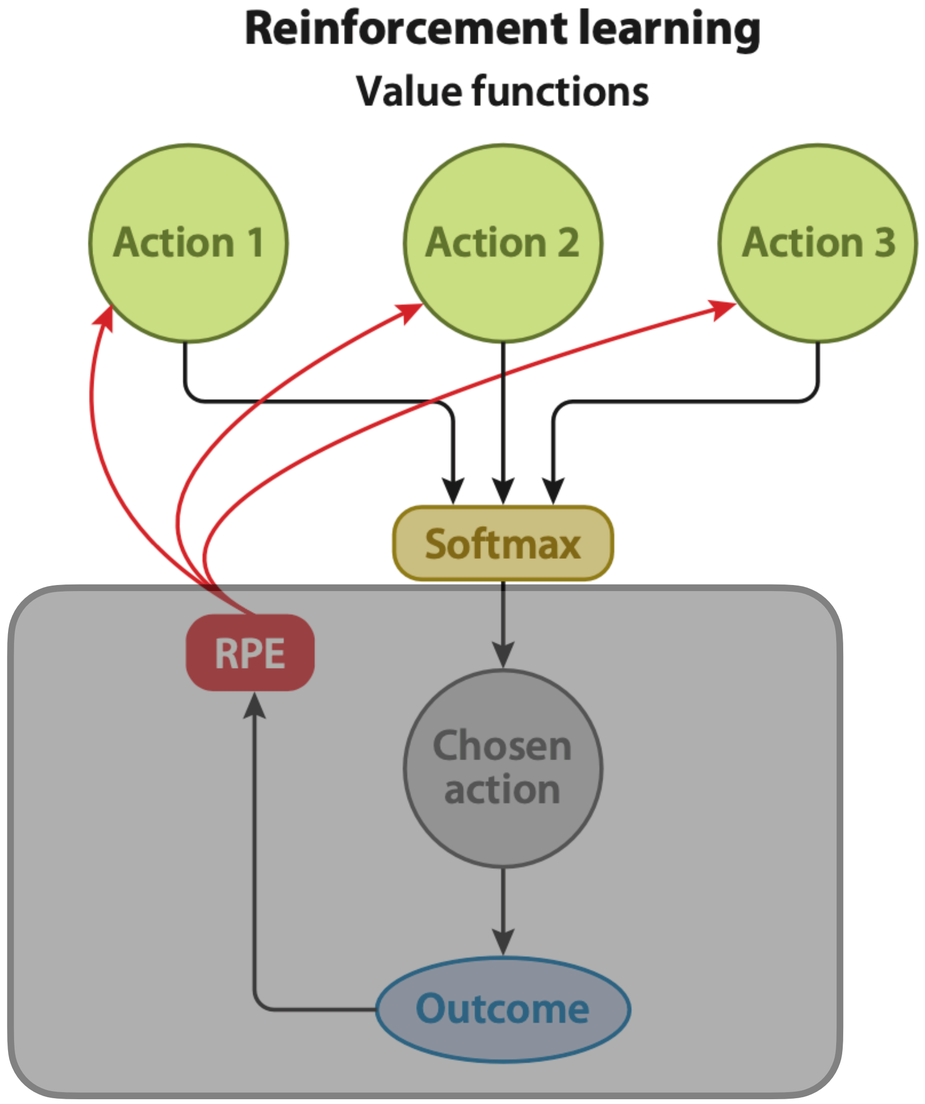


The reward prediction error

Reward prediction error (RPE)

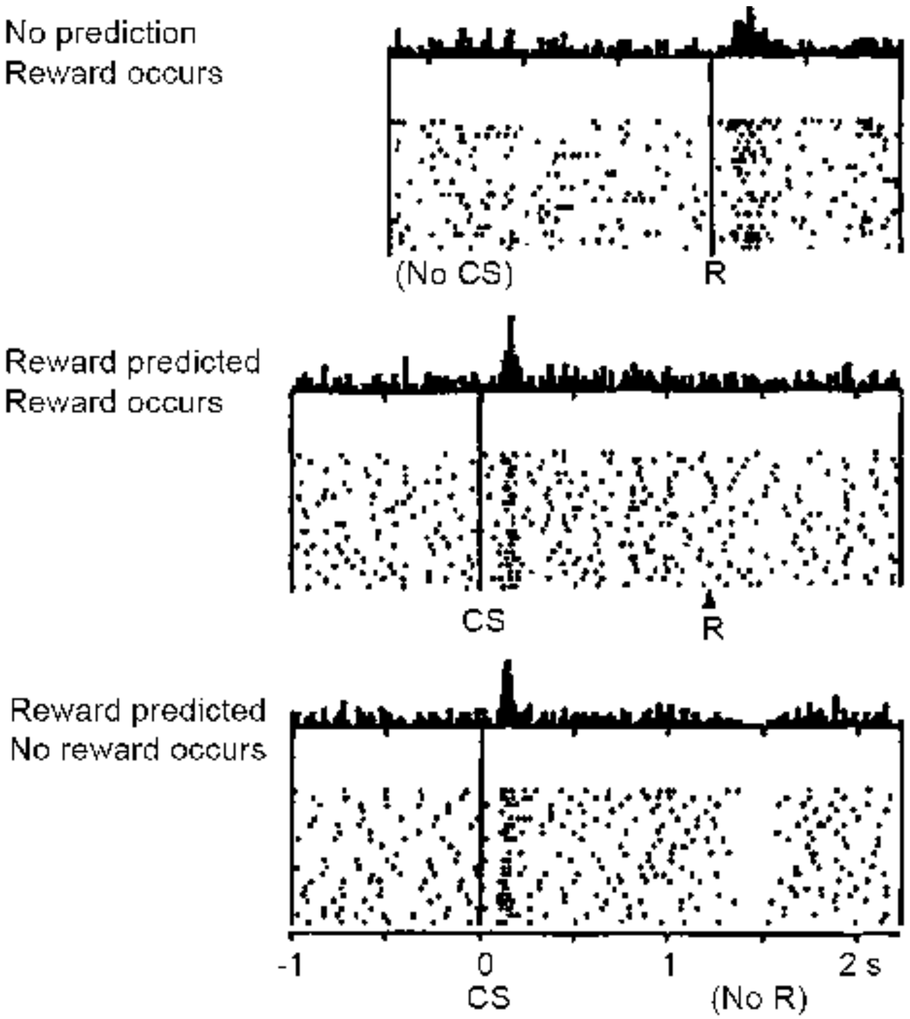
$$Q_{n+1} \doteq Q_n + \alpha \Big[R_n - Q_n \Big]$$

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.

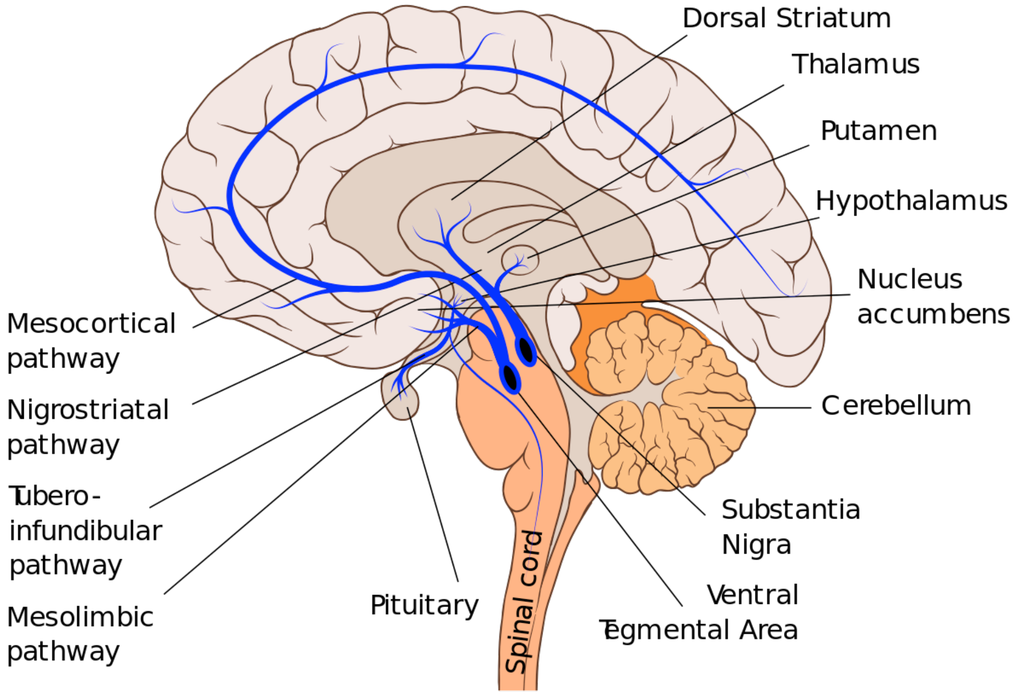


(Lee & Jung 2012)

Phasic dopamine tracks with RPEs



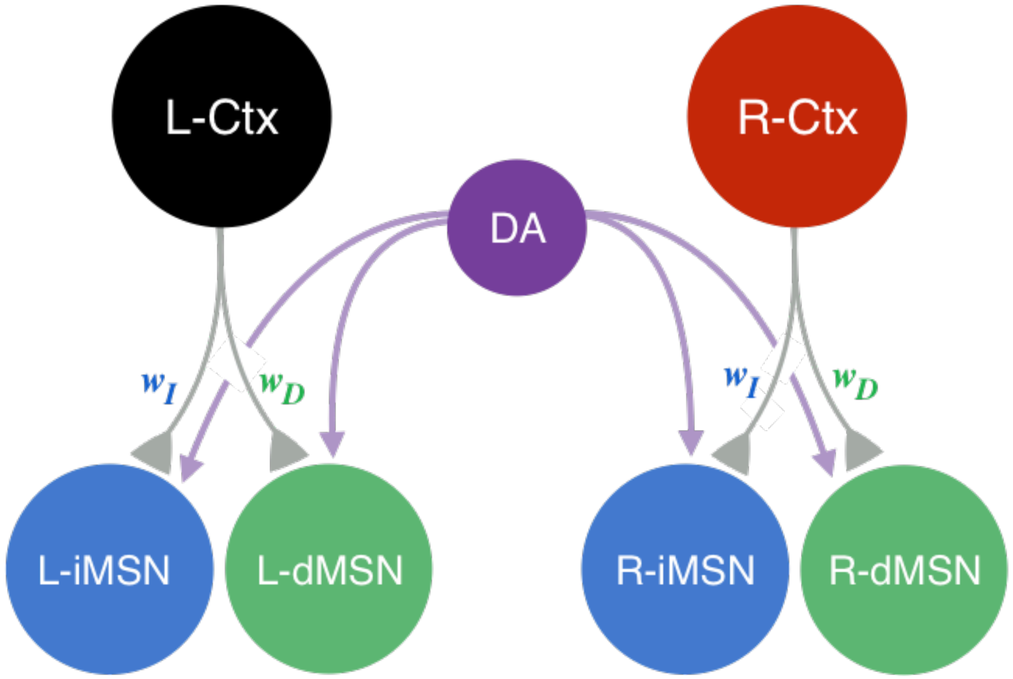
Cells in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) both show activity tied to the reward prediction error.

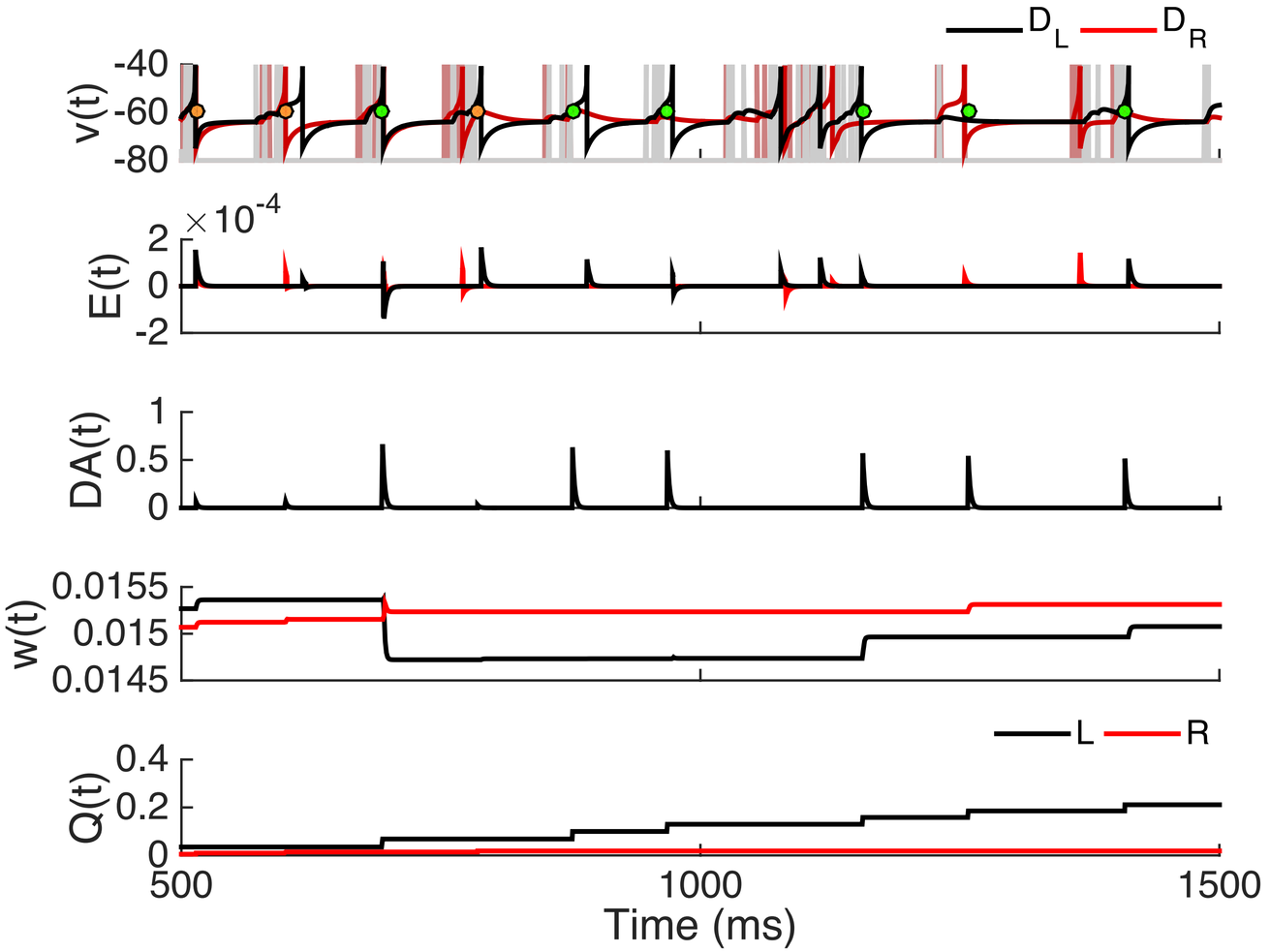


https://en.wikipedia.org/wiki/Dopaminergic_pathways

(Shultz et al. 1997)

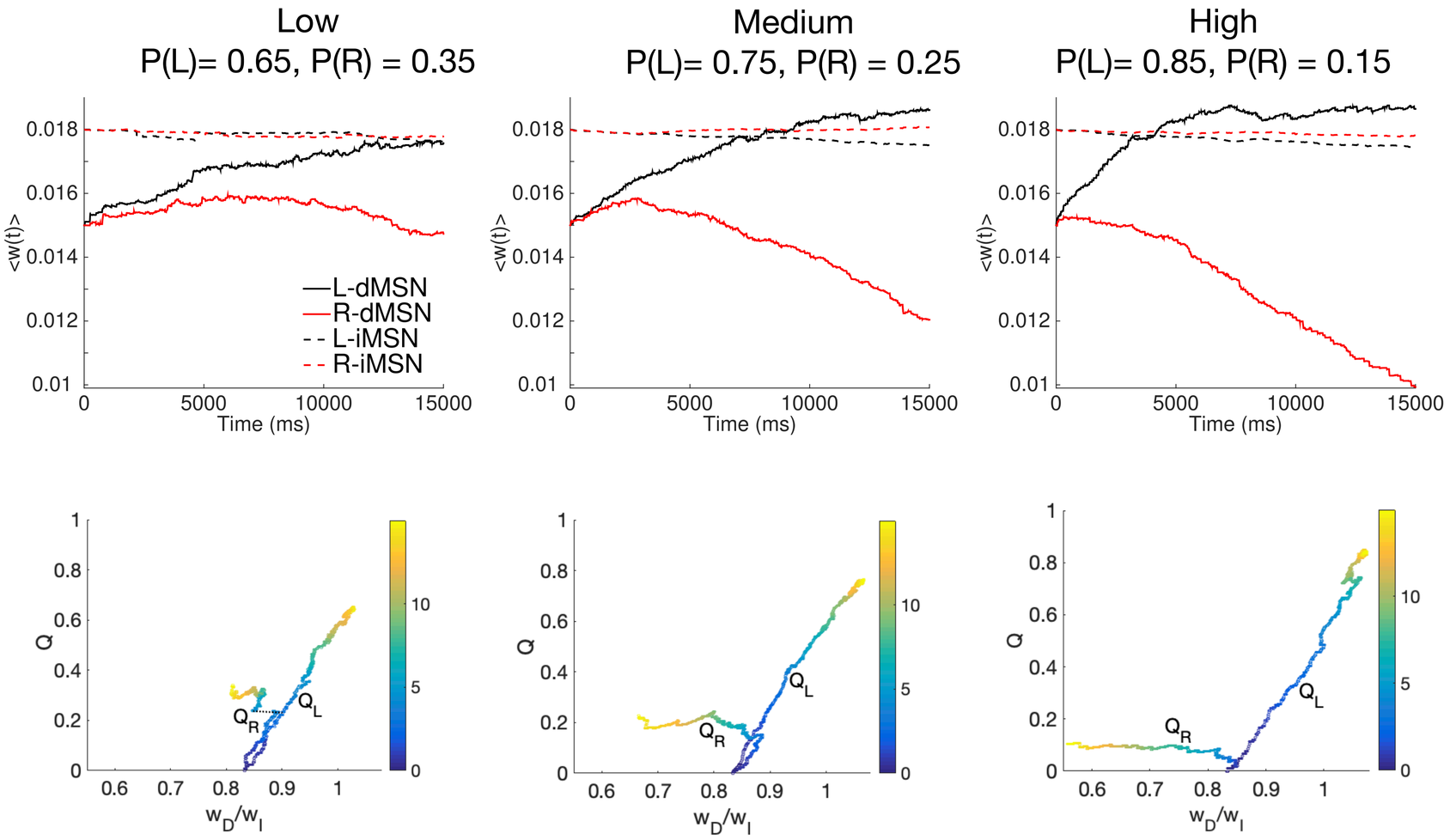
Spike-timing dependent plasticity





(Dunovan et al. 2019; Vich et al. 2020)

Temporal difference update

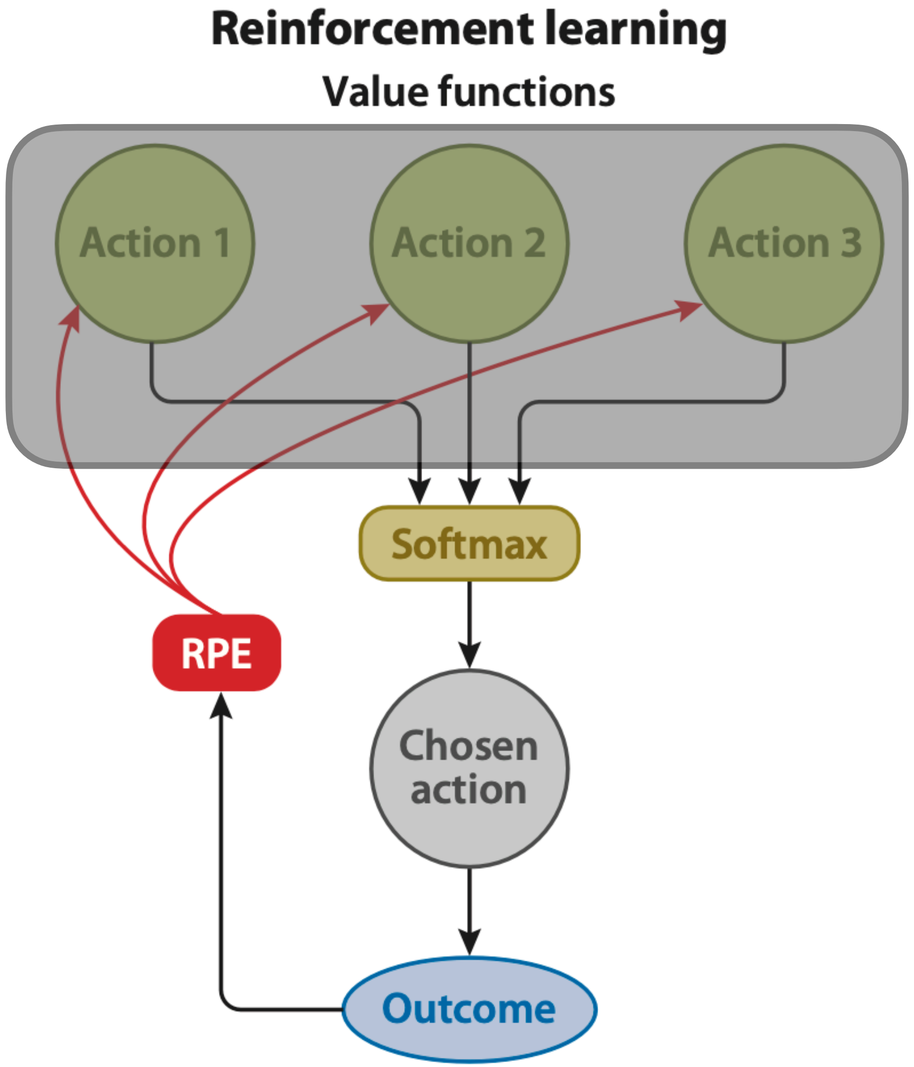


Phasic dopamine signals at the corticostriatal synapses can update behavior consistent with tracking action value (Q)

(Dunovan et al. 2019)

Representation of value

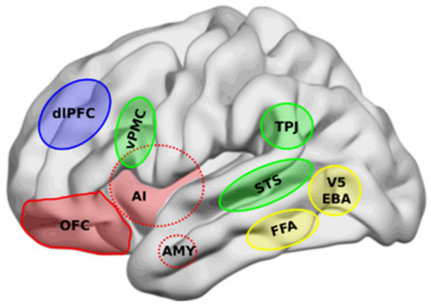
Where is value represented?



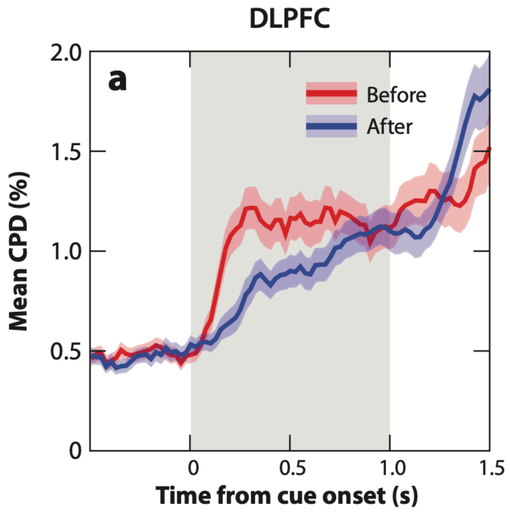
If RPE's drive an adjustment in the value of possible actions, then where does the brain represent (and update) action values?

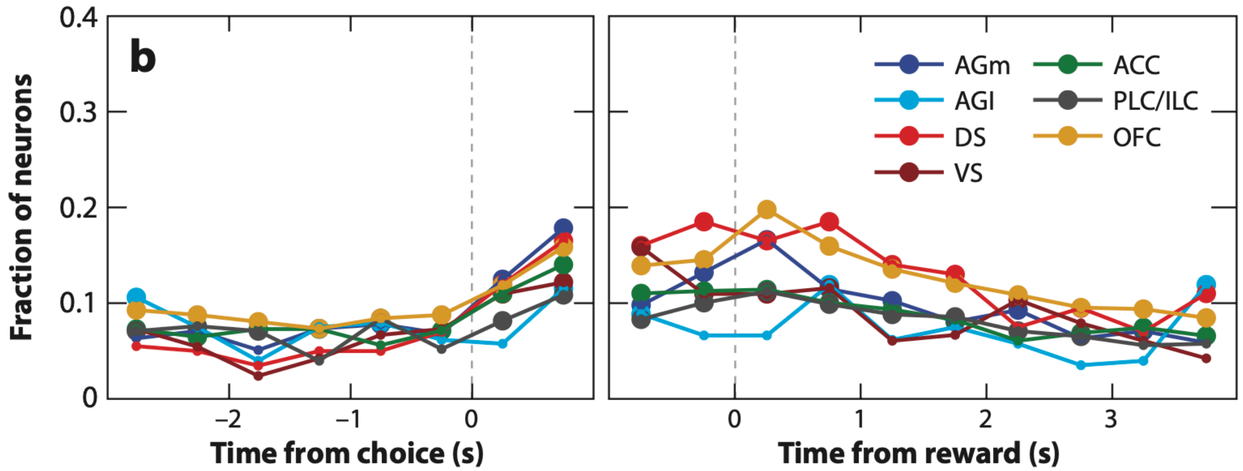
(Lee & Jung 2012)

Neural responses to value after training

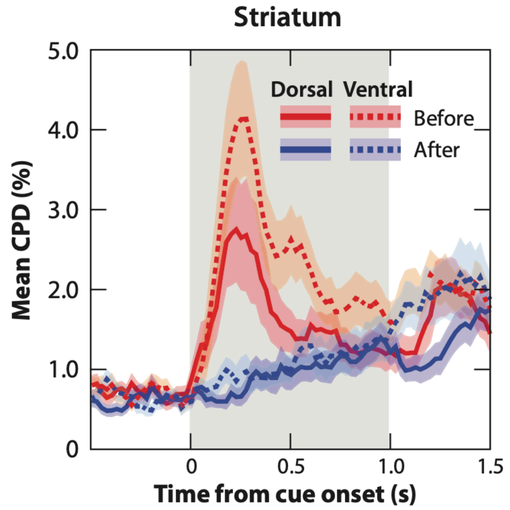


(Credit: Wikipedia)





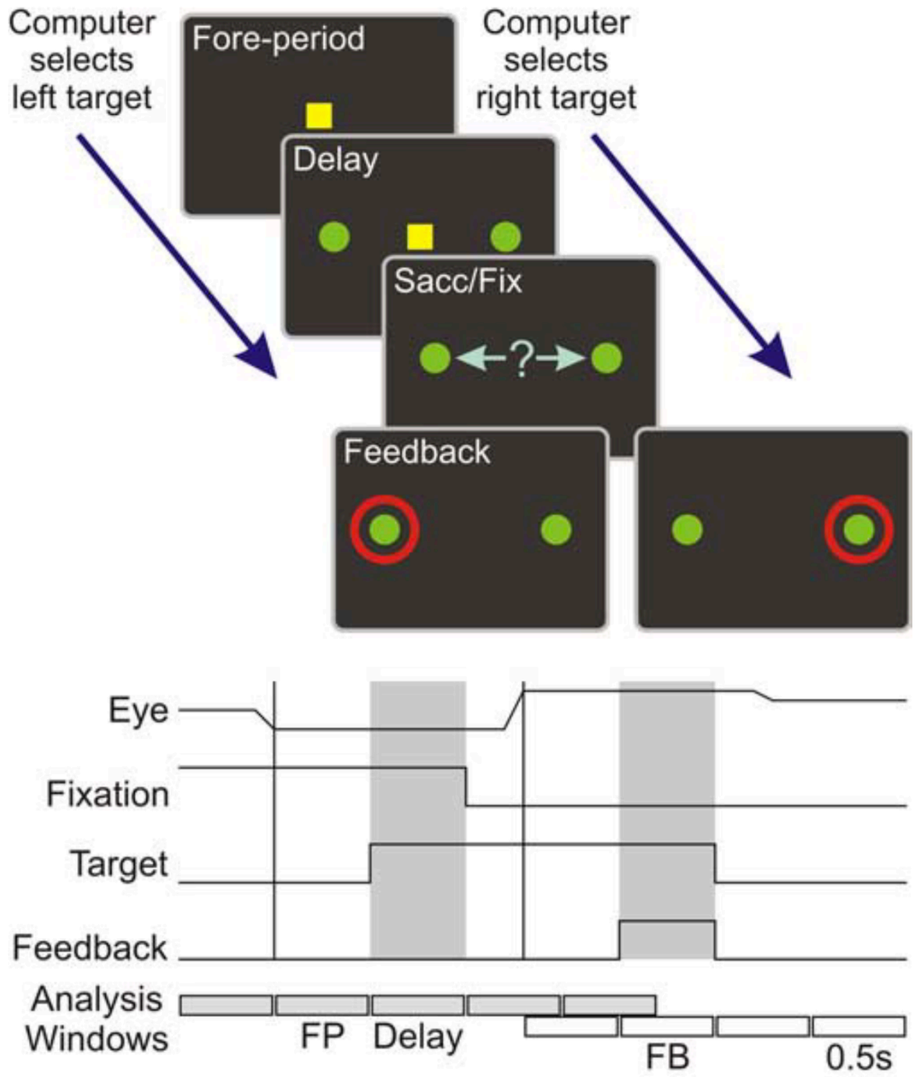
(Credit: Wikipedia)



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.

(Lee & Jung 2012)

The matching pennies task



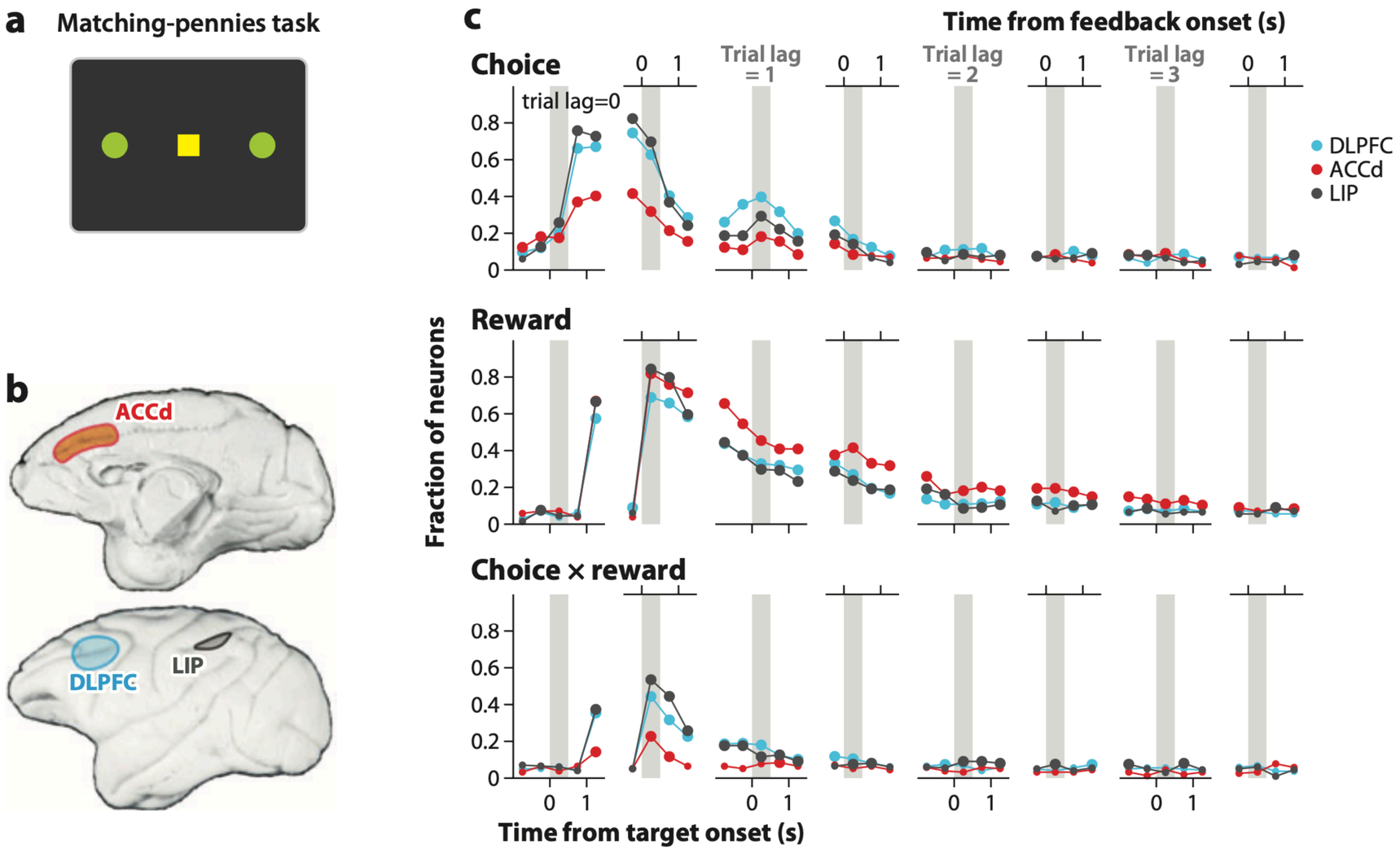
The matching pennies task is a variant of the bandit task where a computer tracks the players performance to estimate their action values to each target and either give feedback consistent with their value (exploitation) or against the value (exploration).

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

(Seo & Lee, 2007)

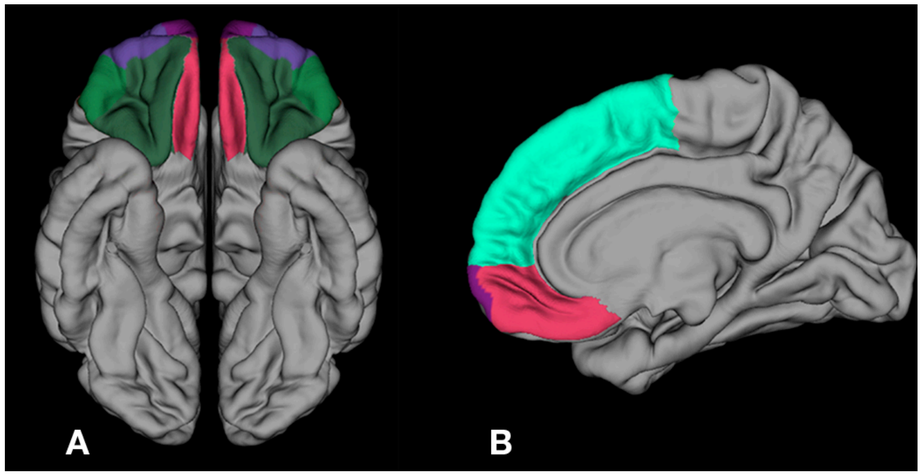
Representing choice, value, or both?

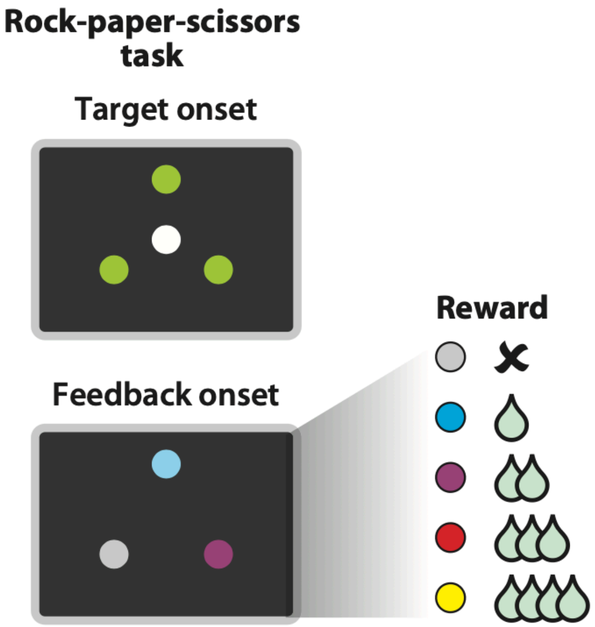


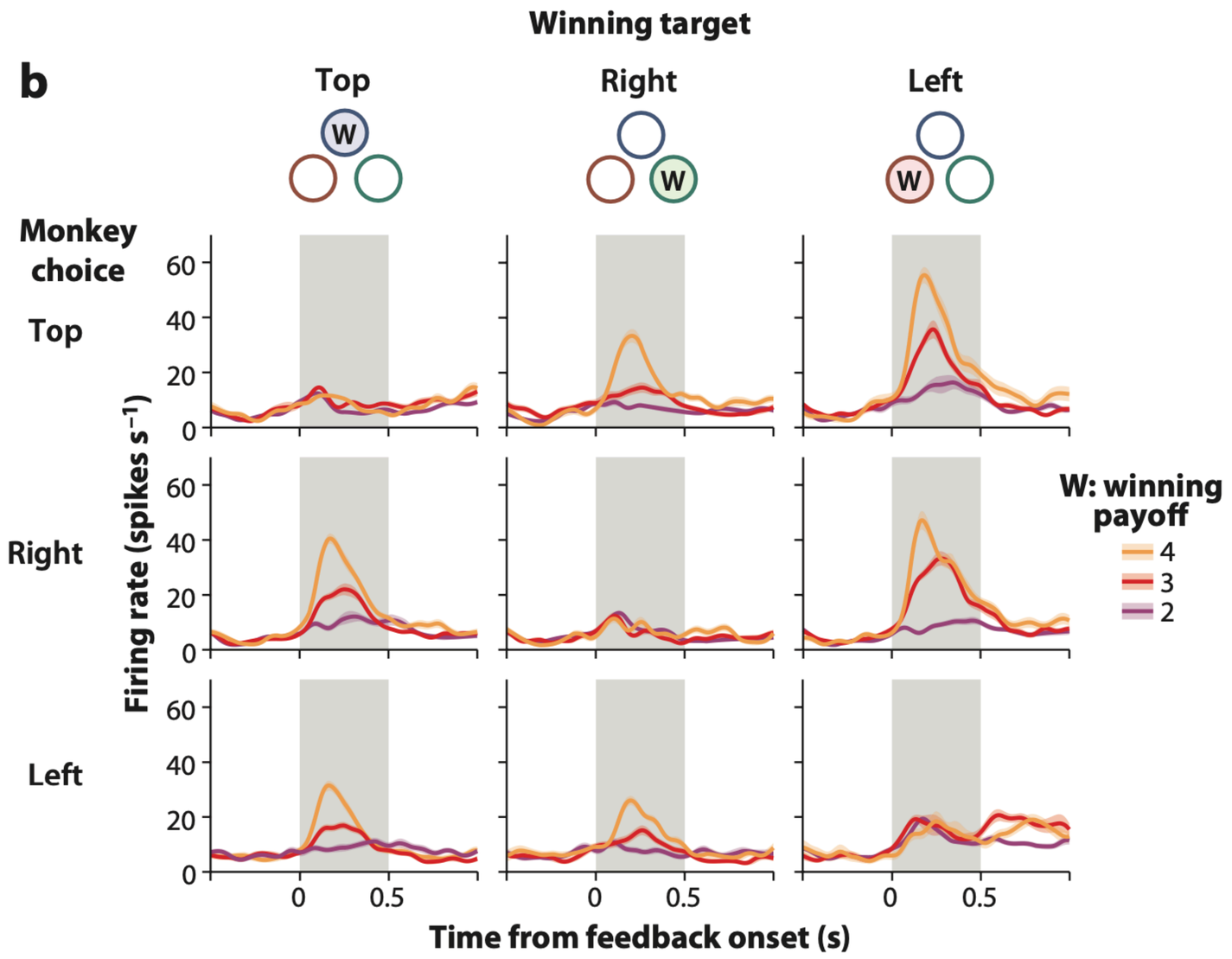
- Lateral prefrontal and parietal regions had a higher proportion of cells tuned to choice.
- Medial prefrontal regions (ACC) had a higher proportion of cells tuned to reward in later trials.

(Lee & Jung 2012)

Orbitofrontal cortex reflects RPEs and value







(Lee & Jung 2012)

Take home message

- Phasic dopamine signals in the substantial 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.

Small group discussions

Answer these two questions:

1. Is dopamine truly acting as a “reward” signal in the brain? Why or why not?

2. In what ways do reinforcement learning models oversimplify the role of dopamine in behavior? Elaborate on additional factors that you think should also be considered.