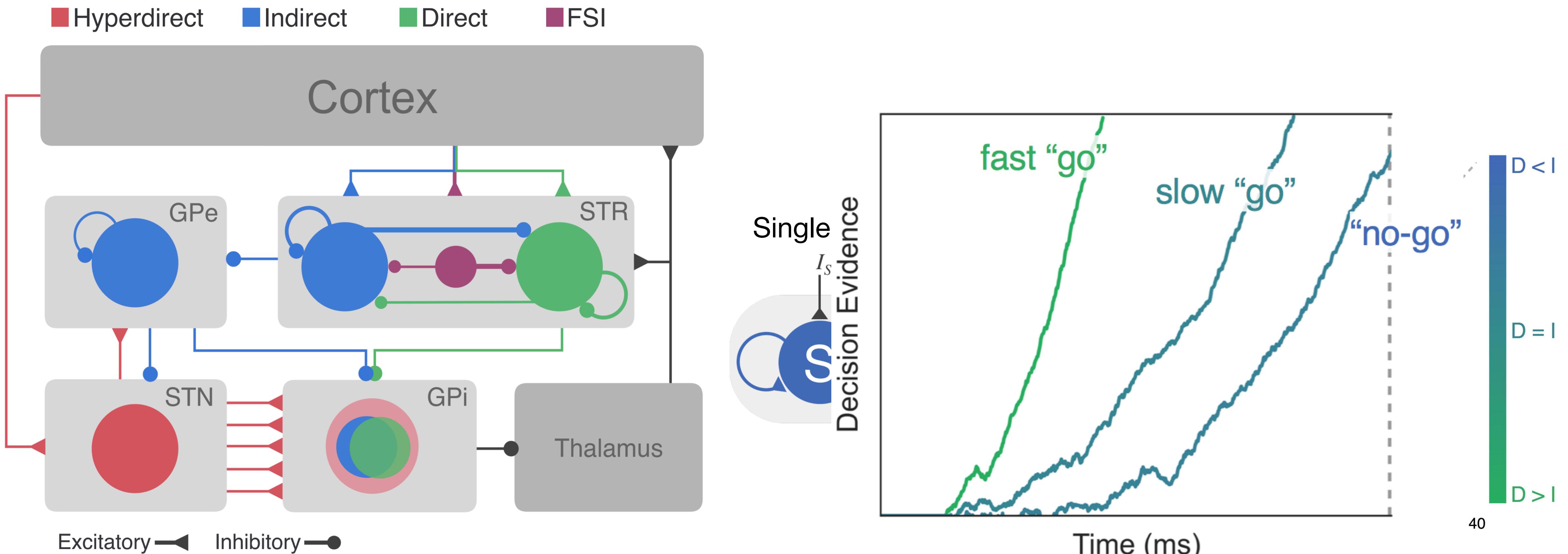


How does the brain learn from feedback?

Readings for today

- Dunovan, K., & Verstynen, T. (2016). Believer-skeptic meets actor-critic: rethinking the role of basal ganglia pathways during decision-making and reinforcement learning. *Frontiers in neuroscience*, 10, 106.
- Dunovan, K., Vich, C., Clapp, M., Verstynen, T., & Rubin, J. (2019). Reward-driven changes in striatal pathway competition shape evidence evaluation in decision-making. *PLoS computational biology*, 15(5), e1006998.

The Believer-Skeptic model



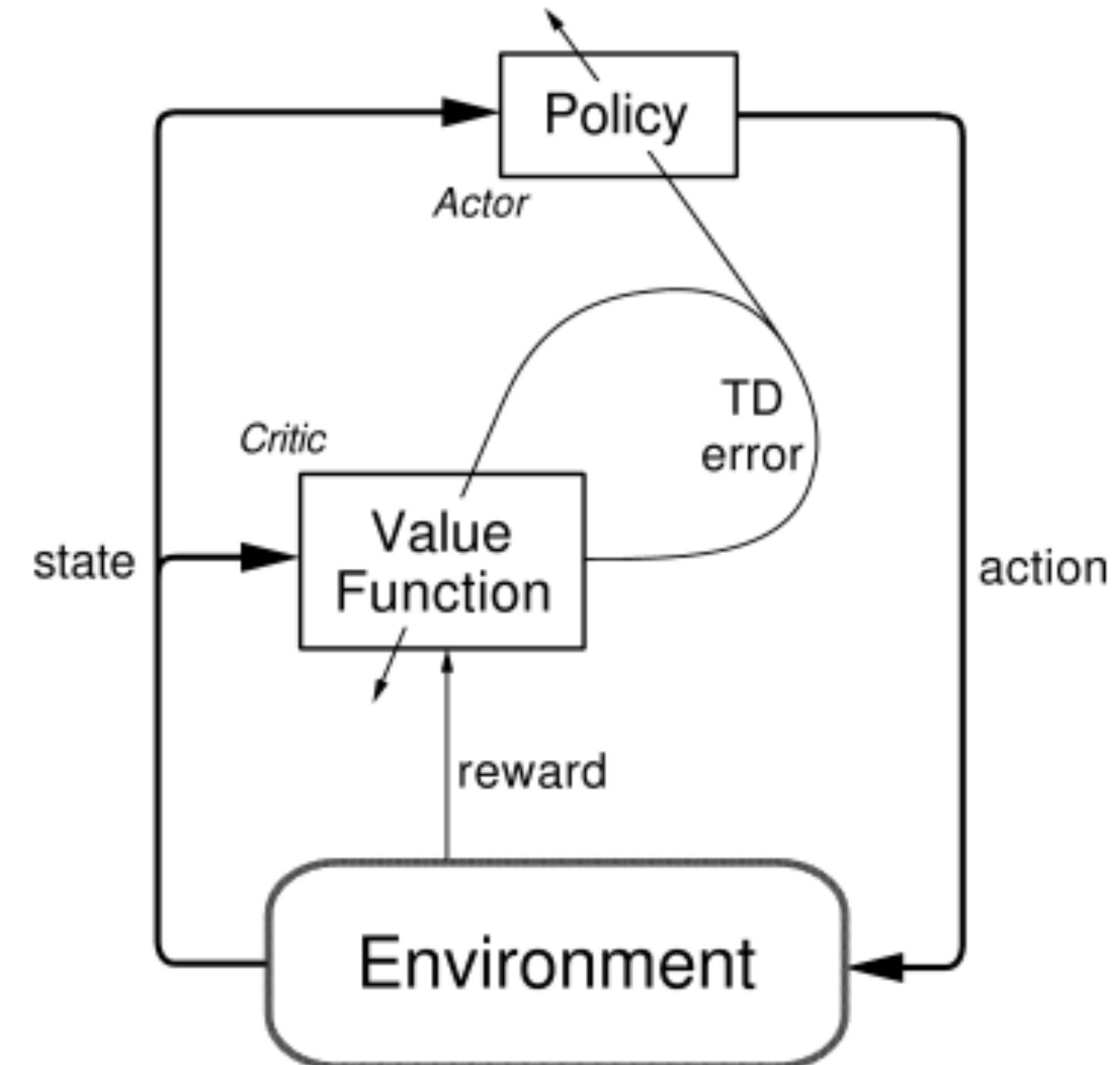
Str: striatum; GPe: external globus pallidus;
GPI: internal globus pallidus; STN: Subthalamic nucleus

Actor-Critic Learning

Two interacting processes (networks):

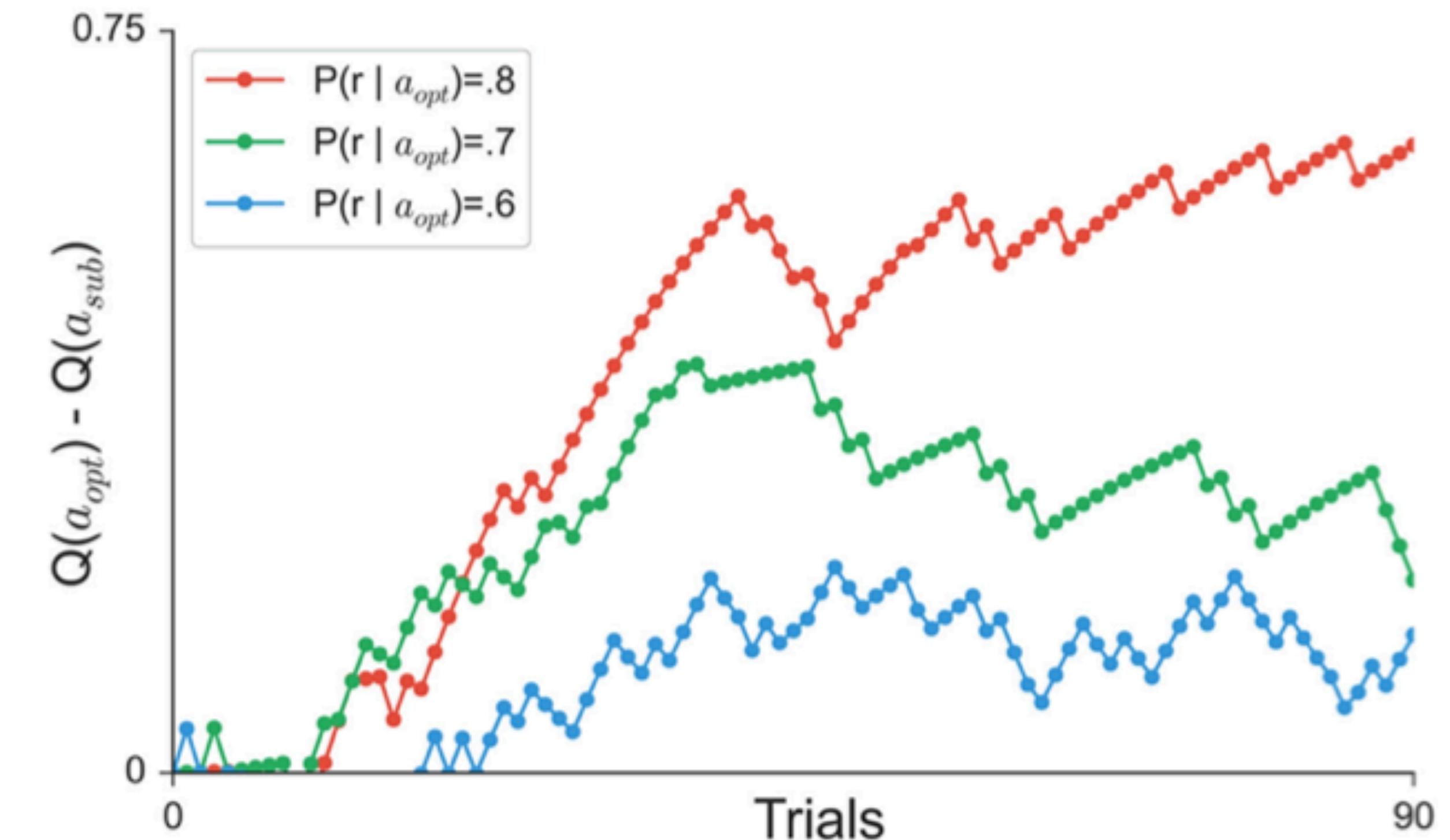
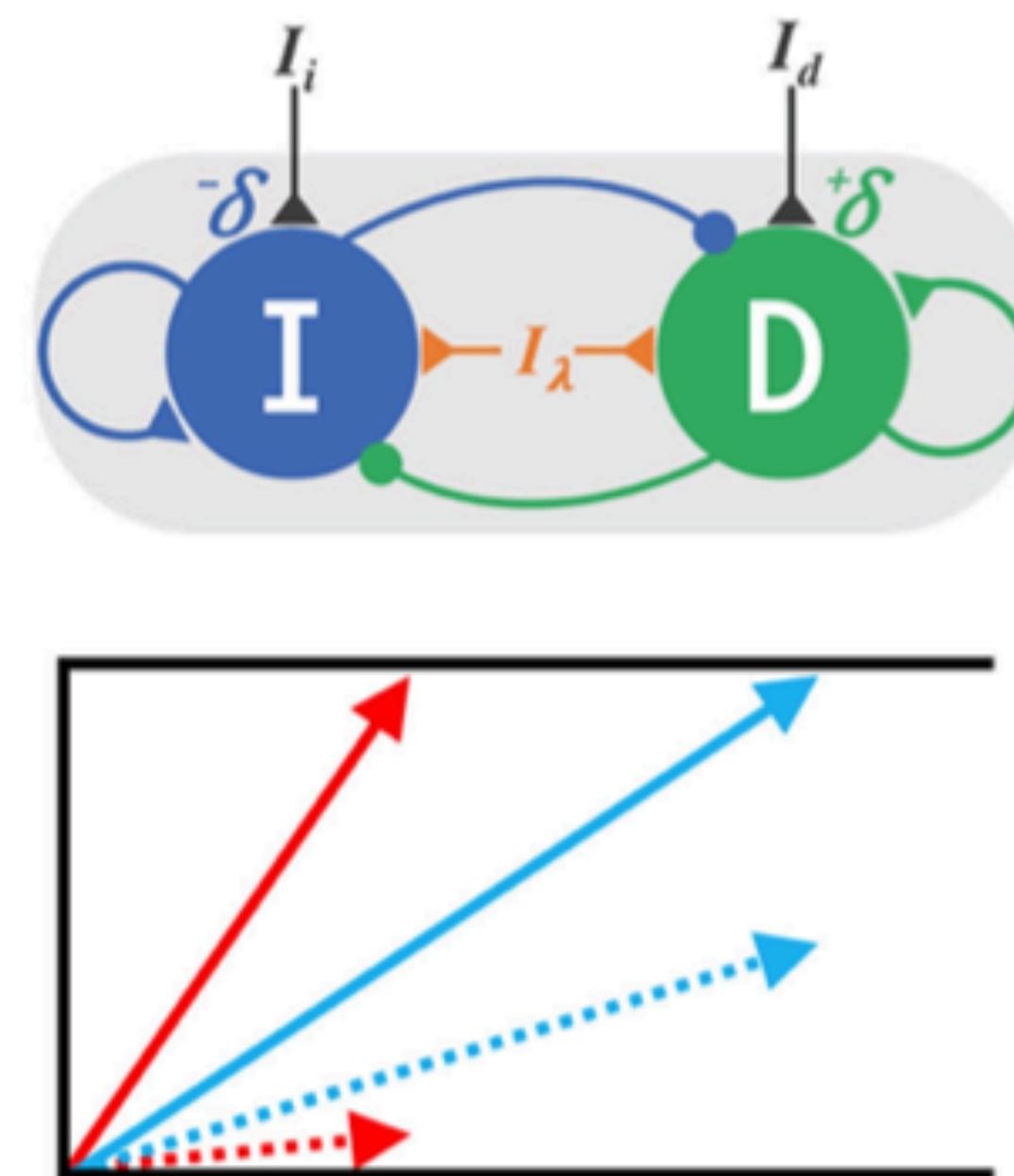
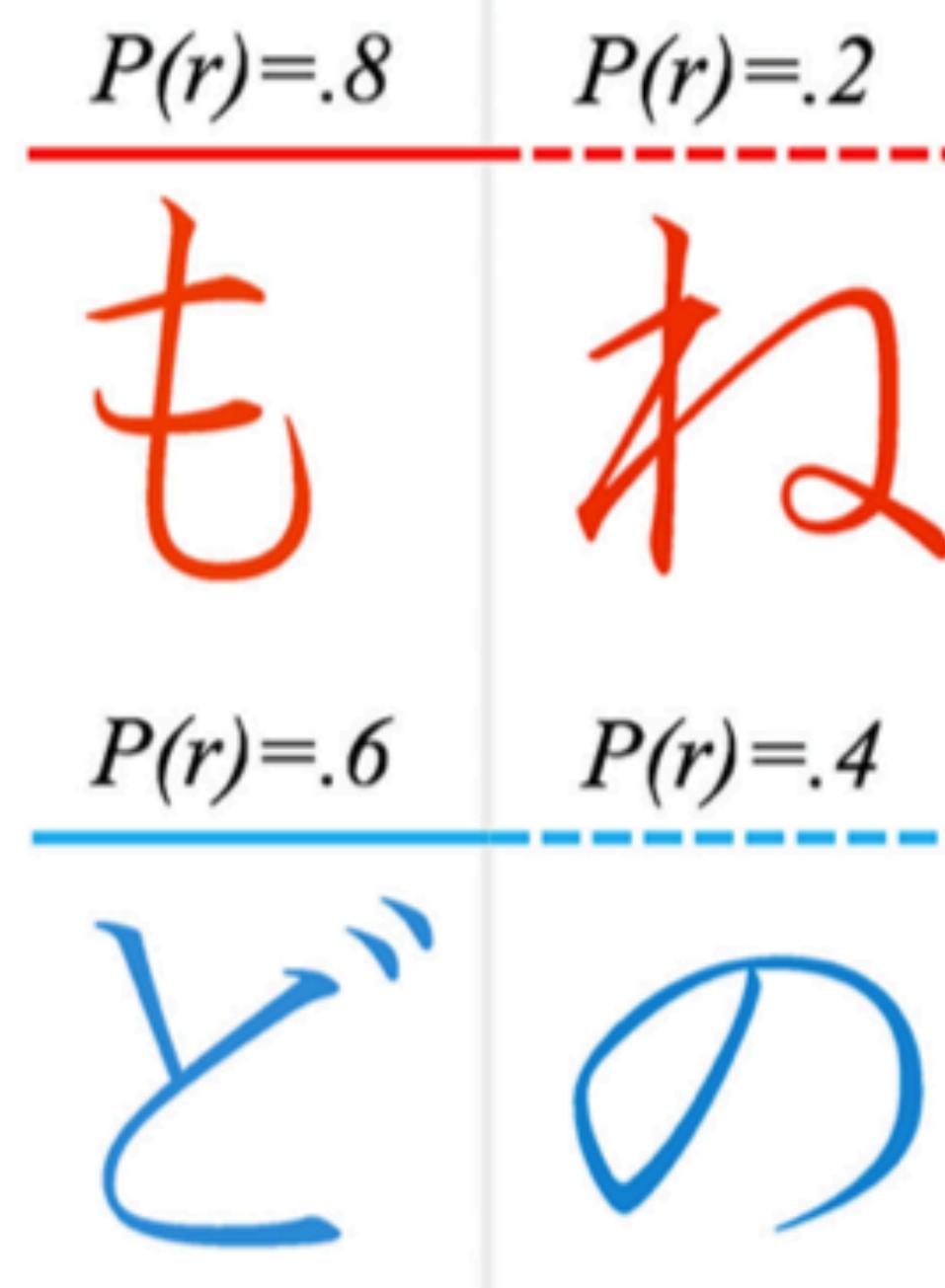
- The actor: decides which action should be taken.
- The critic: informs the actor how good the action was and to adjust adjust.

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

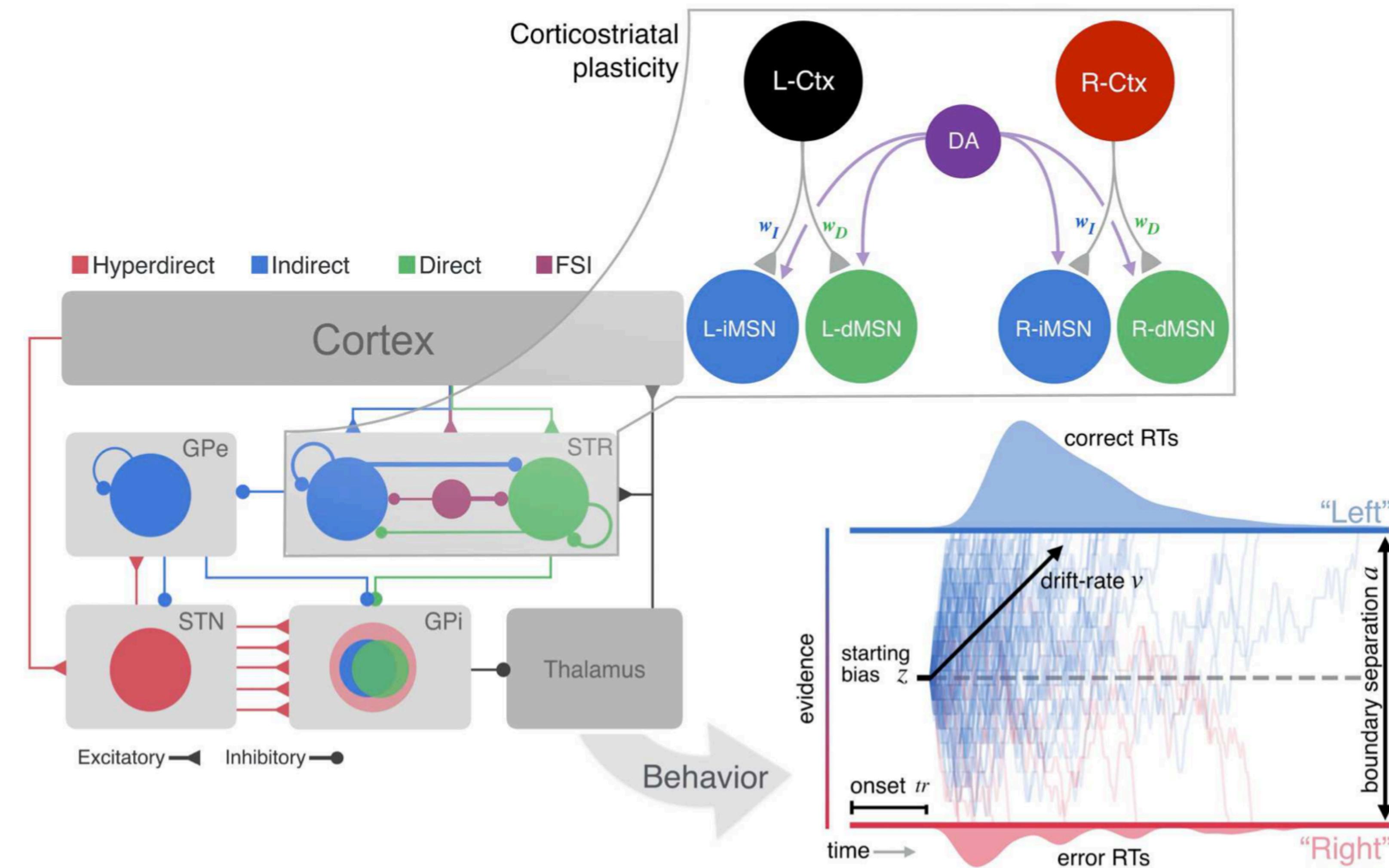


Just add (reinforcement) learning

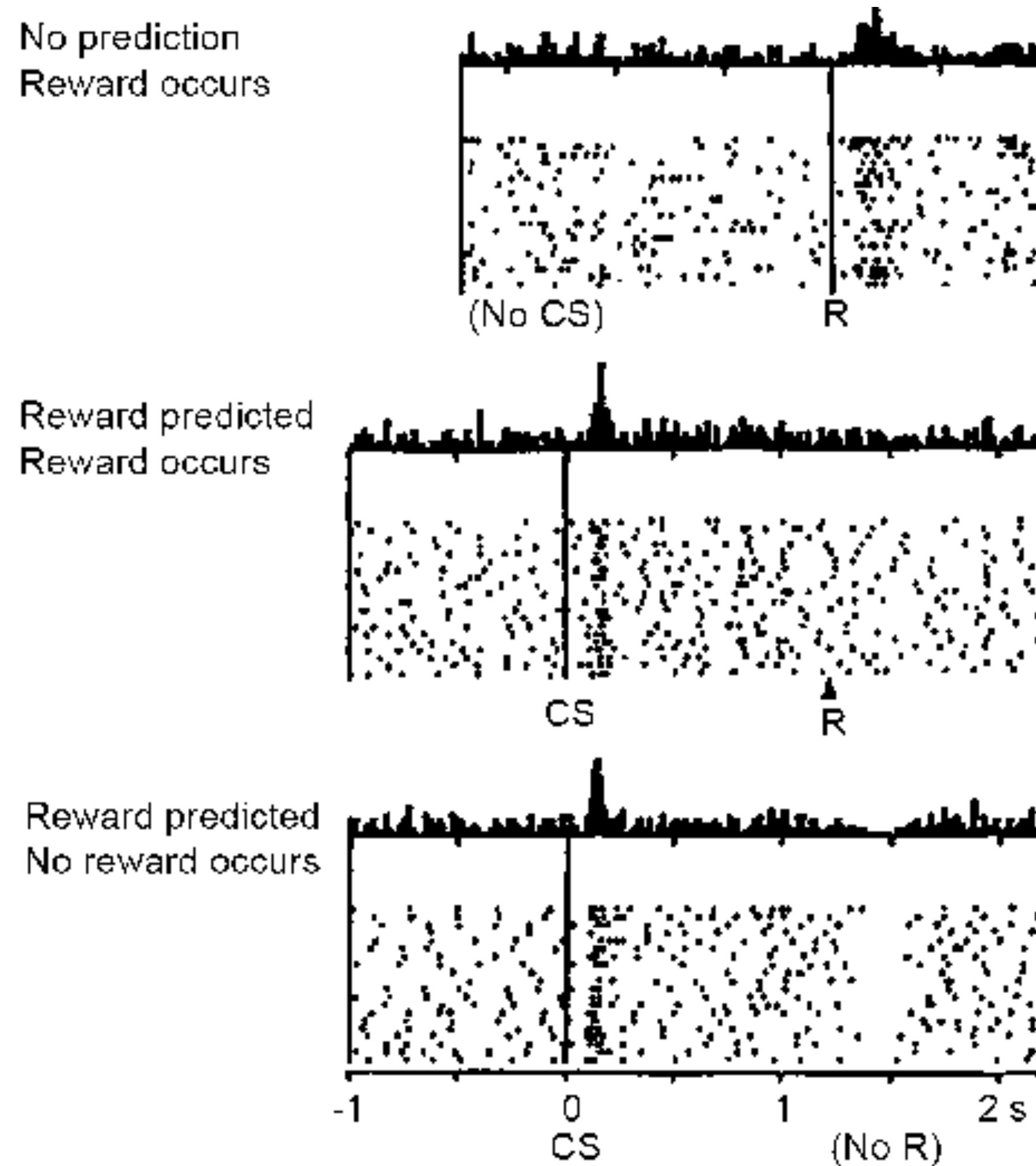
Phasic Dopamine ($\pm \delta$) and RL



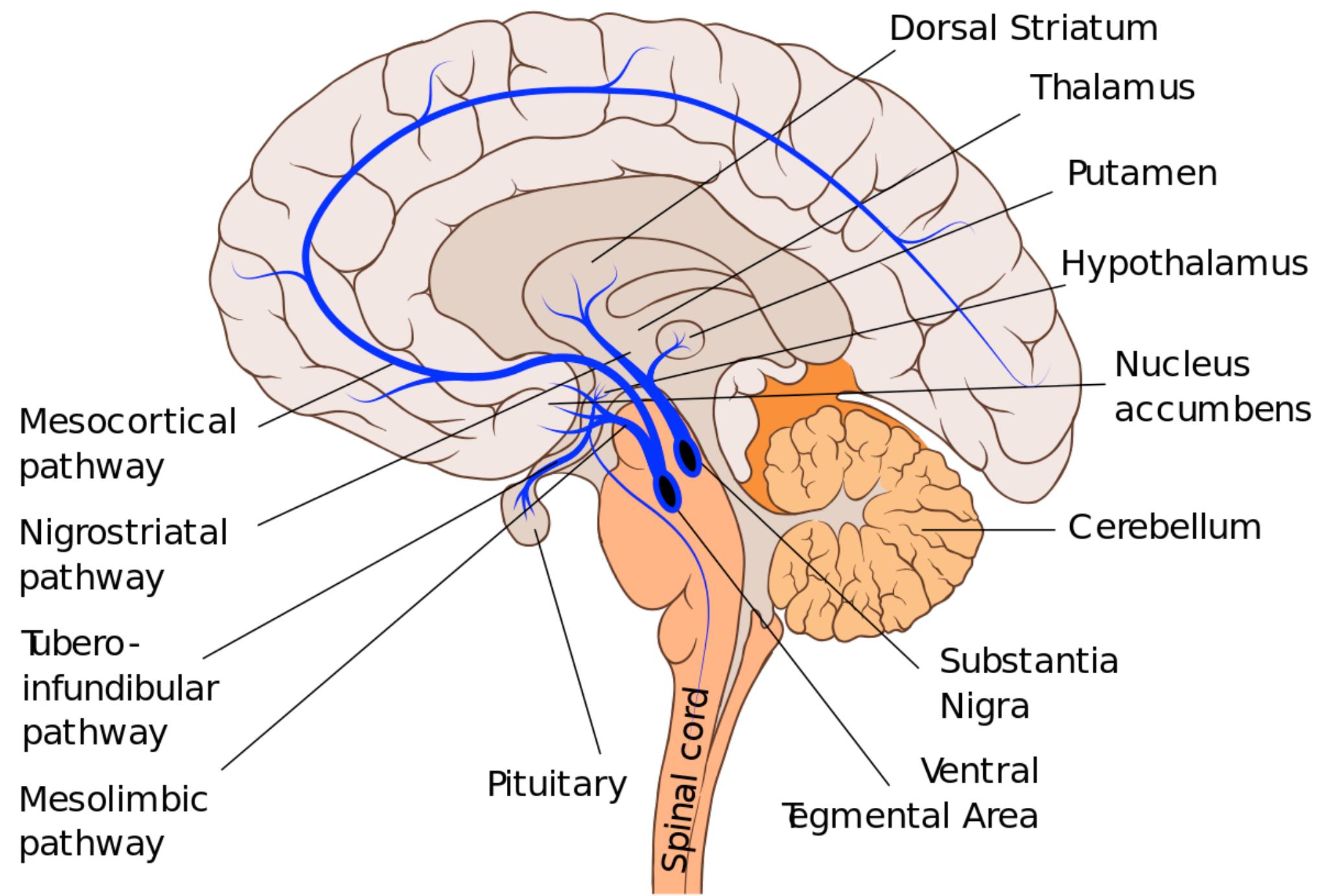
Upwards mapping



Phasic dopamine

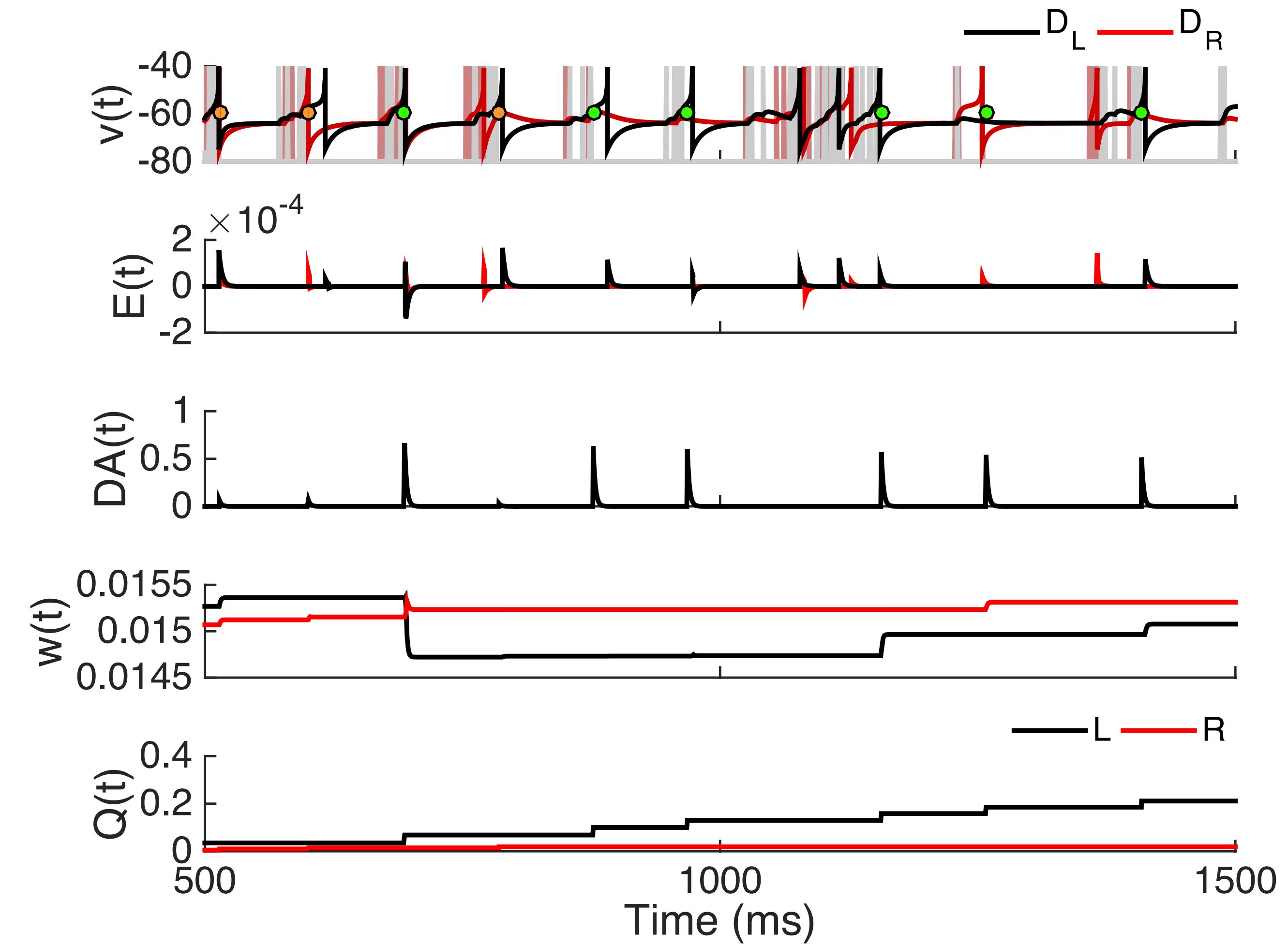
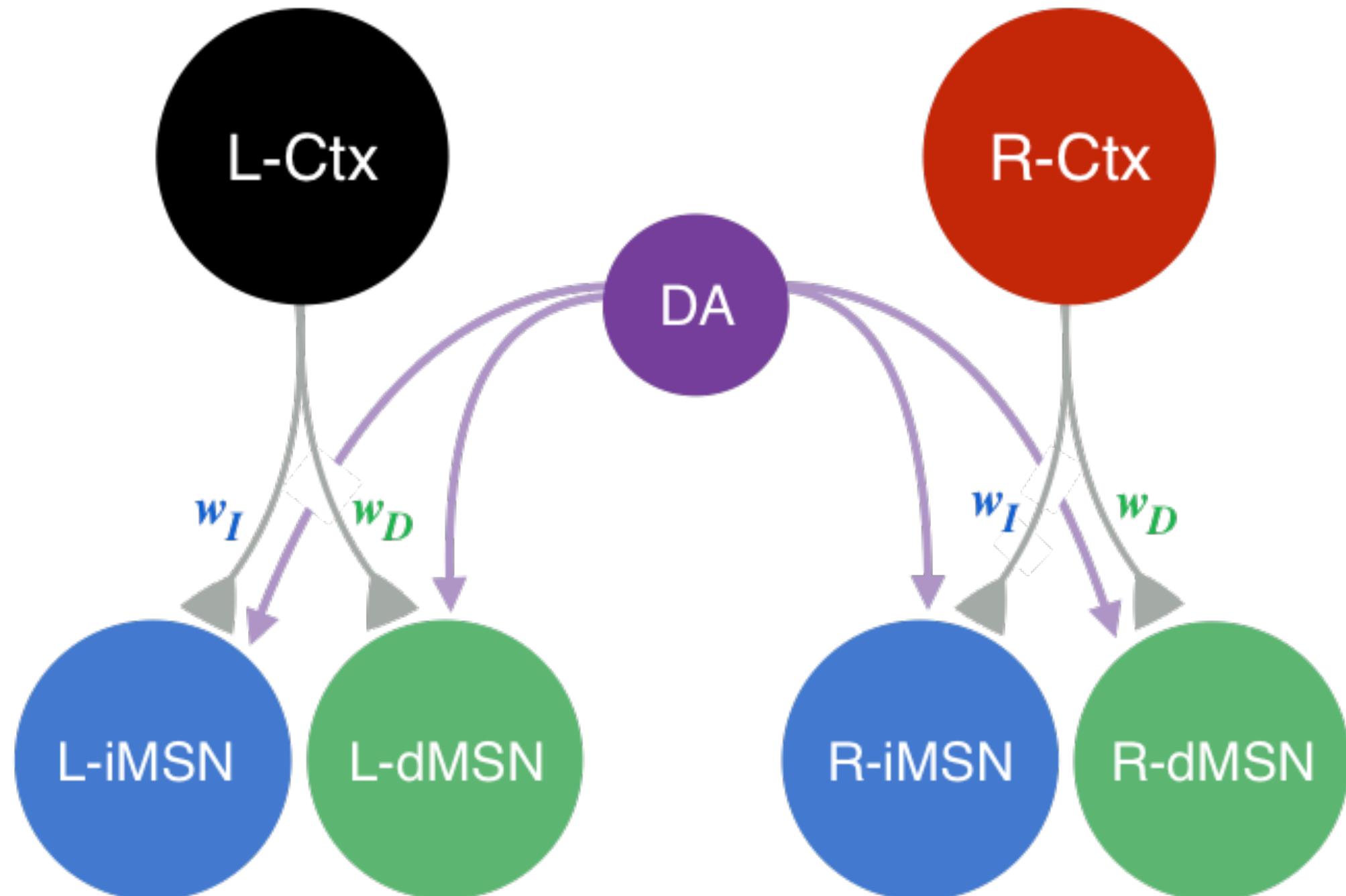


Cells in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) both show activity tied to the reward prediction error.

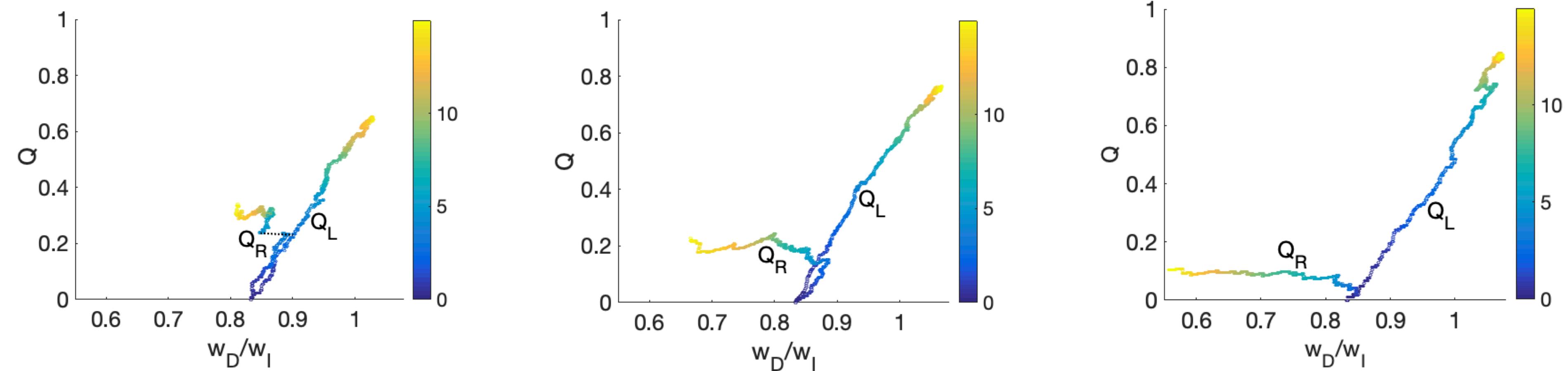
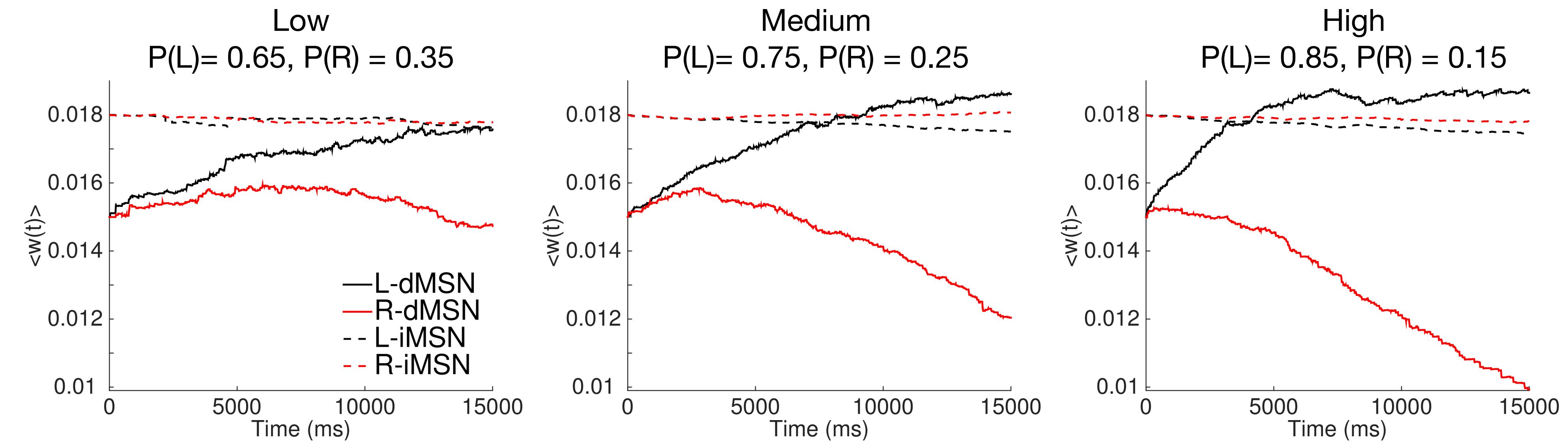


https://en.wikipedia.org/wiki/Dopaminergic_pathways

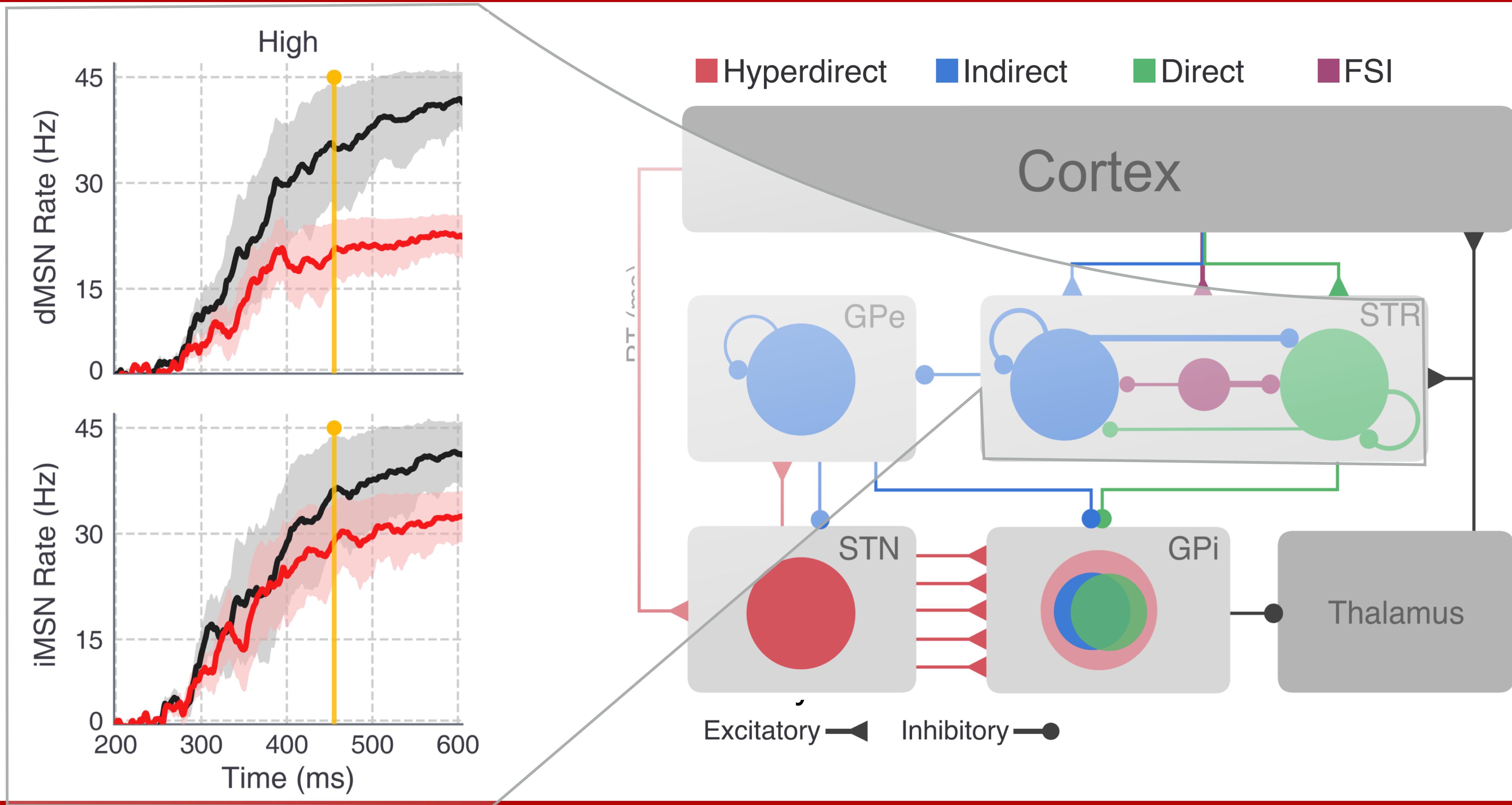
Spike-timing dependent plasticity



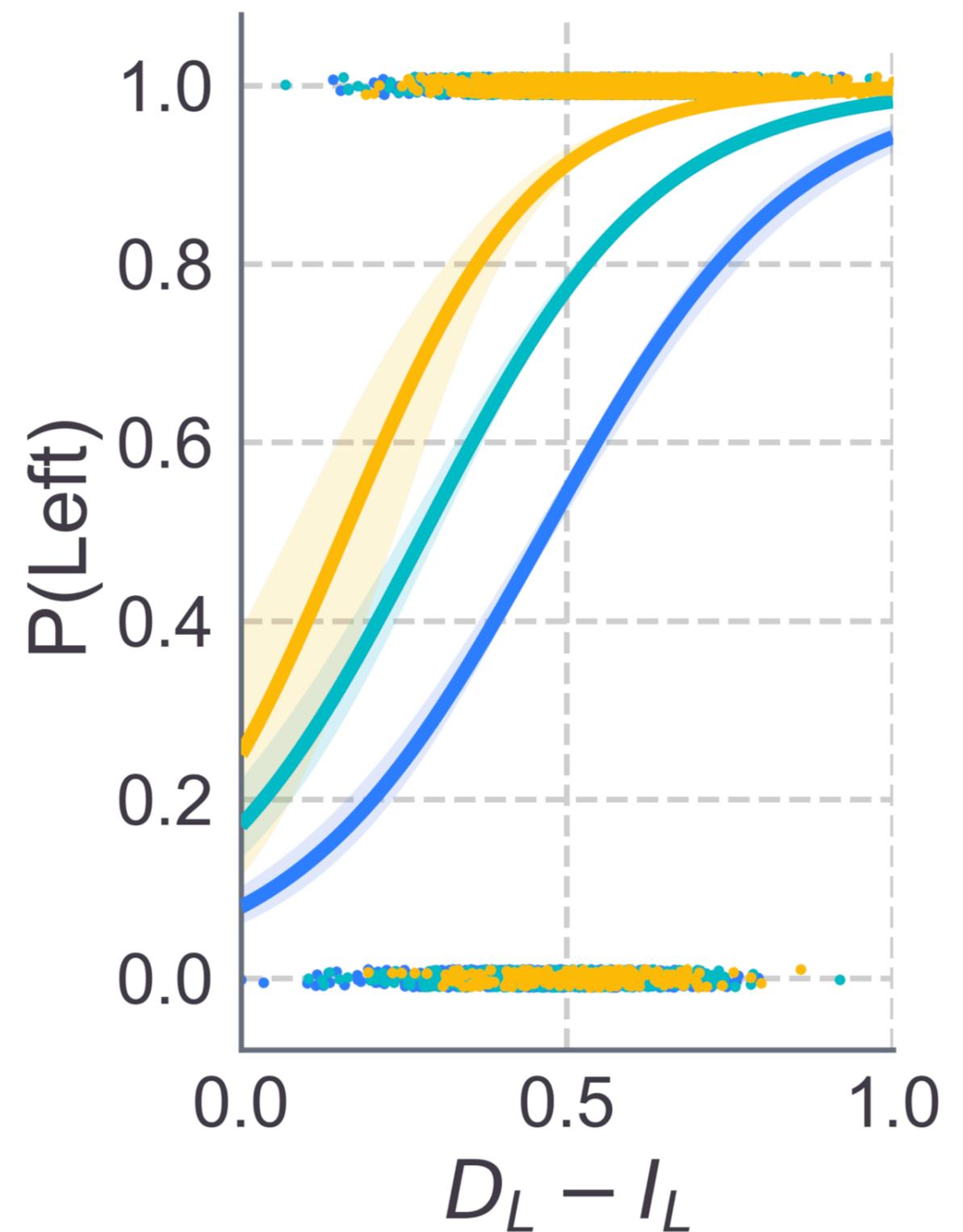
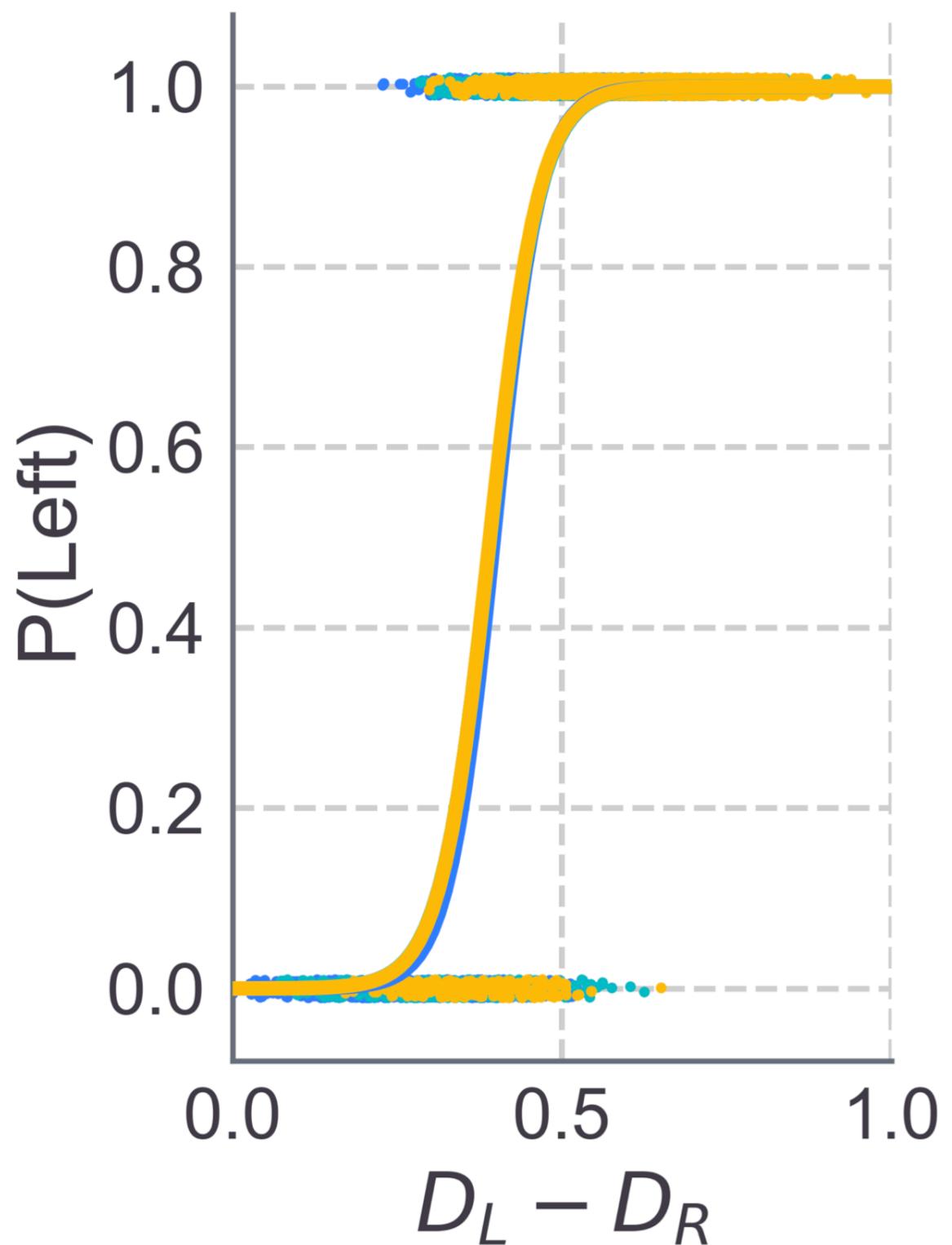
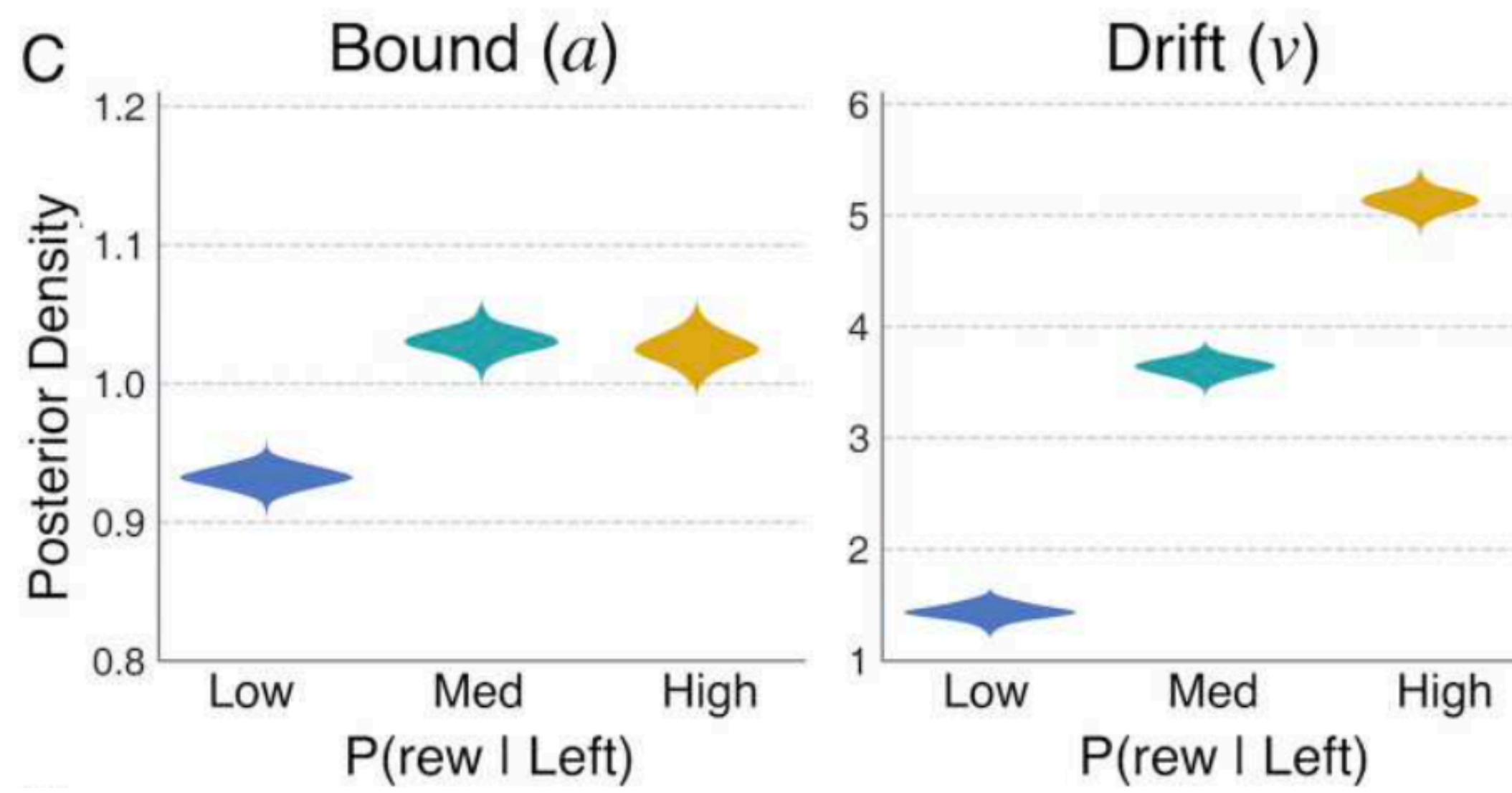
Temporal difference update



Dopamine learning changes B-S balance

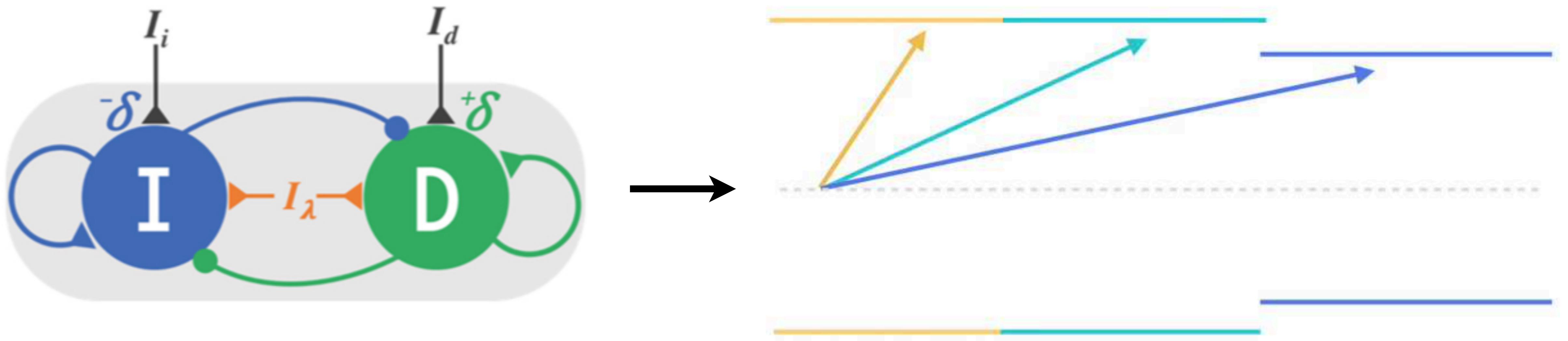


The k -armed bandit problem



Increased separation of direct & indirect pathways (driven largely by changes in direct pathway gain) track with changes in drift rate (v) and action choices.

Believer-Skeptic meets Actor-Critic



Take home message

- Given the role of phasic dopamine in the cortico-basal ganglia circuits, it is easy to see how an Actor-Critic mechanism can be integrated into the Believer-Skeptic model.
- Phasic dopamine, reflecting the RPE, changes the relative balance between Believers (direct) and Skeptics (indirect) pathways to bias the system towards (or away from) actions that lead to more rewards.