Week 7, Mon 5/11

Dan Birman

Schultz et al. set out to connect two observations made in the literature about learning in simple tasks. First, that in behavioral experiments learning is driven by predictions about rewards and punishments. Second, that dopaminergic neurons in primates appear to signal the error relative to the prediction. They argue that a model of temporal difference learning may be sufficient to account for these findings. TD learning was designed based on how animals (appear) to learn, but initially made no explicit predictions about how the learning algorithm should be implemented in a brain. Schultz et al. cite other results showing that physiological recordings from dopaminergic projections appear to show an "error code" that depends crucially on prior predictions. They go on to describe a neurally plausible implementation of TD learning. Finally, they show that a simplified model of their implementation shows response patterns that are qualitatively similar to the recordings made in monkey dopaminergic projections. They suggest therefore that the dopaminergic system (in VTA and sub nigra) may be used by other brain regions as a prediction error signal. Their explanation immediately begs the question of how the brain might implement reinforcement learning on much larger timescales. They hint at possible solutions in the RL literature, but I'm not aware of any physiological recordings that support those claims—would be very cool to hear more about that.