

Temporal difference learning is favored for rewards, but not punishments, in simulations and human behavior

Adam Morris
Brown University

Fiery Cushman
Brown University

Abstract: Evidence indicates that dopaminergic neurons in basal ganglia implement a form of temporal difference (TD) reinforcement learning. Yet, while phasic dopamine levels encode prediction errors of rewarding outcomes, the encoding of punishing outcomes is weaker and less precise. We posit that this asymmetry between reward and punishment reflects functional design. In order to test this hypothesis, we constructed a reinforcement learning algorithm that parameterizes TD learning separately for reward and punishment. We find that the optimal model relies on temporal difference learning for rewards alone. Moreover, this differentiated model provides a significantly better fit to human behavioral data, similarly showing TD learning for rewards more than for punishments. This may be because information about future rewards must shape an earlier sequence of choices, while information about future punishments need only bias the immediately preceding choice.