Three-factor cortico-striatal plasticity shifts activity of cortico-basal ganglia-thalamic subnetworks towards optimal performance in decision-making tasks | Jyotika Bahuguna

Jyotika Bahuguna 1† , Timothy Verstynen 2 and Jonathan E. Rubin 3
1 Laboratoire des Neurosciences cognitive et adaptive (LNCA), Strasbourg, France
2 Department of Psychology, Neuroscience Institute, Carnegie Mellon University, Pittsburgh, USA
3 Department of Mathematics, University of Pittsburgh, Pittsburgh, USA

OVERVIEW: Understanding how cortico-basal ganglia-thalamic (CBGT) circuits influence decision making remains a challenge, especially considering the different decision policies a biological agent can adopt in response to environmental changes and the complexity of interacting pathways in CBGT networks. We deconstruct the process of value-based learning in a CBGT network into three aspects: (a) defining what is a decision policy, (b) identifying where in the CBGT network these decision policies are effectively generated, and (c) analyzing how the CBGT pathways encode and modulate different aspects of decision policies. We use an evidence accumulation model (the drift diffusion model; DDM) to map the behavioral features (e.g., decision times, choices) of the decision making agent into a decision policy (what). Based on our prior work [1], we identified three low dimensional CBGT subnetworks called control ensembles (responsiveness, pliancy and choice) that represent control over distinct dimensions of the decision policy (where). We study how CBGT networks modulate decision policies by simulating learning via dopaminergic signals acting on the cortico-striatal projections in a model CBGT network performing a simple two-choice task with one optimal (i.e rewarded) target.

RESULTS: While our naive model CBGT networks lay in an exploration regime, we observed that value-based learning breaks the speed-accuracy tradeoff and drives the CBGT networks in a direction of maximal increase in reward rate such that they arrive at an exploitation regime and approach the optimal performance curve (OPC). The OPC is a theoretical estimation of the normalized decision times that maximize the reward rate as a function of rate of error in the context

of the DDM [2]. This approach towards the OPC was also recently observed in rats performing a perceptual learning task, where the decision times are slower than predicted by OPCs during initial phases of learning, but move towards the OPC as learning progresses [3]. Our use of a model network allowed us to generate predictions about the contributions of the CBGT control ensembles to this process: our results suggest that learning induces an increase in responsiveness (shorter evidence accumulation onset delays), increase in choice (higher rate of evidence accumulation), and decrease in activity of the pliancy components (corresponding to heightened decision boundary). On the shorter timescale of consecutive trials, each possible set of reward outcomes induces a specific adjustment of control ensembles. Interestingly, experiencing at least one unrewarded outcome within two initial trials can lead to faster convergence towards the OPC than that which results from pairs of rewarded outcomes. Overall, our results suggest that dopamine-dependent plasticity in the corticostriatal projection may be a possible mechanism to achieve average reward rate maximization by promoting changes in activity that ripple down through the CBGT network to achieve the coordinated tuning of the activity of its decision policy control ensembles.

[1] Vich C, Clapp M, Rubin EJ & Verstynen T (2022), PloS Computational Biology
[2] Bogacz R, Brown E, Moehlis J, Holmes P & Cohen J (2006) Physiological review
[3] Masis J, Chapman T, Rhee JY, Cox DD & Saxe AM (2023), eLife