Modelling dopamine dynamics: encoding predicted reward in the striatum enables adaptive decision-making within a spiking CBGT network

Alex O'Hare | University of California

Authors: Alex O'Hare (1, 2), Catalina Vich (1, 2), Jonathan E. Rubin (3, 4), Timothy Verstynen (3, 5)

Affiliations: (1) Dept. de Matemàtiques i Informàtica, Universitat de les Illes Balears, Palma, Illes Balears, Spain, (2) Institute of Applied Computing and Community Code, Palma, Illes Balears, Spain, (3) Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania, United States of America, (4) Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, (5) Department of Psychology & Neuroscience Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America

Abstract:
The cortico-basal-ganglia-thalamic (CBGT) pathways are widely held to be responsible for reinforcement learning within vertebrates. Nigrostriatal dopaminergic projections have been shown to signal the value of reward prediction error, and facilitate learning via plastic changes in the synaptic weights of spiny projection neurons (SPNs) in the striatum. Recent studies have proposed that the predicted reward (Q-value) may be reflected in the difference between the corticostriatal synaptic weights of the D1 and D2 receptor-expressing SPNs. Building on this idea, in this work we develop a model of dopaminergic learning and incorporate it into a fully spiking neural network of the CBGT pathways.

We investigate the extent to which this model may enable adaptive decision-making under a multi-armed bandit task and resolution of the explore-exploit dilemma by incorporating an uncertainty bonus which would align with experimental observations from humans and non-human primates. In this work, we develop a mathematical model representing the amount of DA made available to the SPNs, P, such that

$$dP/dt = (DA\_T - P)(s + (r-Q))-bP,$$

where 'DA_T' denotes the total pool of available DA, 's' is a baseline level of DA, 'b' is a decay rate in the amount of available DA, 'r' represents the reward value, and 'Q' the predicted reward. The Q-value is determined by an effect size measure given by

$$Q = 2*(W^{dSPN}-W^{iSPN})/(N^{dSPN}+N^{iSPN}),$$

where W^dSPN and W^iSPN represent the synaptic weights of the D1-expressing direct pathway SPNs (dSPNs) and D2-expressing indirect pathway SPNs (iSPNs), respectively. N^dSPN and N^iSPN represent the number of dSPNs and iSPNs involved in the transaction, respectively. Preliminary simulations with a toy model (see supporting document) have yielded promising results, with the resulting Q-value aligning closely with the reward probability implemented in the trials.

We explore how our model supports recent studies which show the lack of distinction between tonic and phasic DA signals and instead produce a DA signal which aligns more closely with that of a value-function. Through our investigation we aim to reconcile this conflict between the traditional and emerging views of the characteristic features of dopamine within the decision-making pathways of the basal ganglia.

In demonstrating the efficacy of this model in a spiking neural network, we will lend support to the idea that the Q-value may be encoded in the synaptic weights of the SPNs. Furthermore, by removing the need to explicitly code a Q-value update rule into the network, we enhance the automaticity and biological realism of the existing open-source framework, CBGTPy, which may be used to conduct computational experiments of decision-making tasks.
