Reinforcement Learning by a Biologically Constrained Spiking Neural Model of the Basal Ganglia

Carlos Enrique Gutierrez[1*], Jean Lienard[1], Benoît Girard[2], Kenji Doya[1]

[1]*Neural Computation Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan*
[2]*Institut des Systèms Intelligents et de Robotique, Université Sorbonne, Paris, France*

While evidence suggests that reinforcement learning (RL) is realized by the basal ganglia (BG), it is still unknown how computations like stochastic action-selection and chosen action value update are realized by the spike communications and neuromodulation in the BG. We constructed a BG spiking neural network model based on anatomical constraints and multi-objective parameter search to match physiological data [Lienard&Girard 2014, Girard et al. 2018, Gutierrez et al. 2018]. Here we test how the BG model in an arm-reaching task with 3 targets can realize RL. The model includes cortical areas (S1,M1,M2), the BG (striatum,GPe,STN,GPi), the thalamus, a simplified arm model, and reward feedback by dopamine.
Target positions are coded by neural populations in M2 and M2 output to the striatum cause a cascade of spike responses in GPe, STN, and GPi. Action-selection is realized when a sub-population of GPi suppressed, which activates sub-populations in the thalamus and M1. The M1-cerebellar circuit then generates the arm movement trajectory with the hand position coded in S1. Upon hand reaching to the preferred target, dopamine is released over all the cortical synapses to the striatum. Synapses incorporate dopamine-dependent STDP at excitatory connections. The responses of striatal sub-populations change over the time, similarly to the action values update in RL.
We ran episodes using NEST 2.18 [Jordan et al. 2019]. We observed that chaotic dynamics within the BG model allows GPi to select a target randomly during the early episodes and that eventually the preferred channel is selected with the relevant synapses reinforced.

Keywords: reinforcement learning, basal ganglia, dopamine-dependent STDP, action-selection, spiking neural network model

Email:carlos.gutierrez@oist.jp