Resolving the credit assignment problem in cortico-basal ganglia pathways

Matthew Clapp[1,2], Jonathan Rubin[2,3], Catalina Vich[4], Kendra Noneman[5], Timothy Verstynen[2,6]

[1] Neuroscience Inst., Carnegie Mellon University. [2] Center for the Neural Basis of Cognition, Carnegie Mellon University & University of Pittsburgh. [3] Dept. of Mathematics, University of Pittsburgh. [4] Dept. de Matemàtiques i Informàtica, Universitat de les Illes Balears.
[5] Micron School of Materials Science and Engineering, Boise State University. [6] Dept. of Psychology, Carnegie Mellon University.

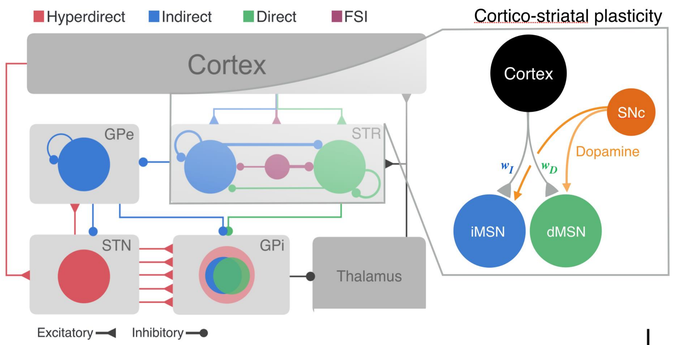
BACKGROUND & MOTIVATION

The cortico-basal ganglia-thalamus (CBGT) network is thought to serve a critical role in learning- and decision-related behaviors.

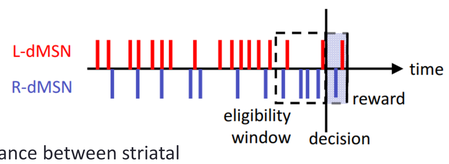
Striatal subpopulations serve as branchpoint between two pathways.
- D1 MSNs form action-facilitating direct pathway
- D2 MSNs form action-suppressing indirect pathway

Cortico-striatal connections, which exhibit dopamine (DA) induced plasticity, strongly influence balance of these pathways and thus overall CBGT network dynamics.

Prediction errors are encoded as phasic dopamine bursts (+PE) or pauses (-PE)
- Cortico-D1 connections potentiated by +PE, depressed by -PE
- Cortico-D2 connections depressed by +PE, potentiated by -PE



Credit assignment problem: how to ensure connections responsible for selected action experience plasticity?

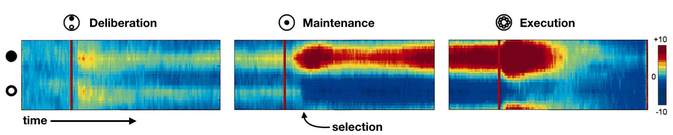


Eligibility traces alone fail due to distance between striatal dynamics, action selection, and reward.

METHODS

Possible resolution: Sustained cortical activity in selected cortical population acts to increase and maintain eligibility in corresponding striatal populations.

Cortical activity known to be sustained in motor planning tasks[7].



Fully spiking basal ganglia network model:
- Two action channels, one per alternative, each containing cortical, basal ganglia, and thalamic populations
- Two striatal populations (D1 and D2) per channel

Sensory input represented by excitatory input to cortical populations

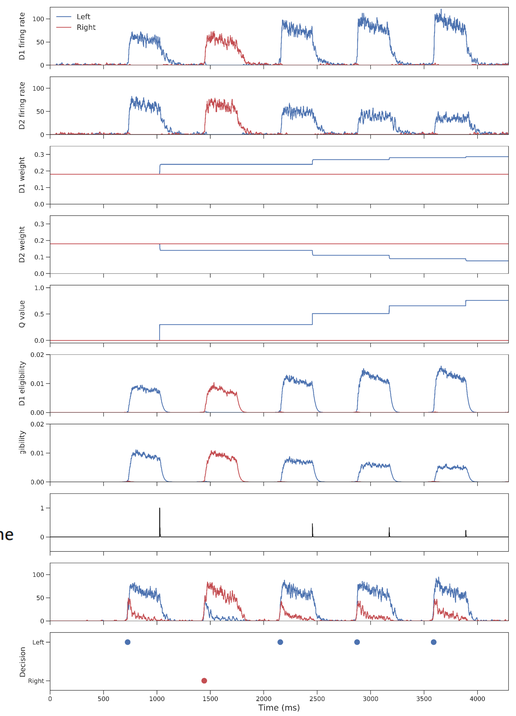
Ramping of thalamic activity to a cut-off (30 spikes / sec) was interpreted as a decision made by the network.

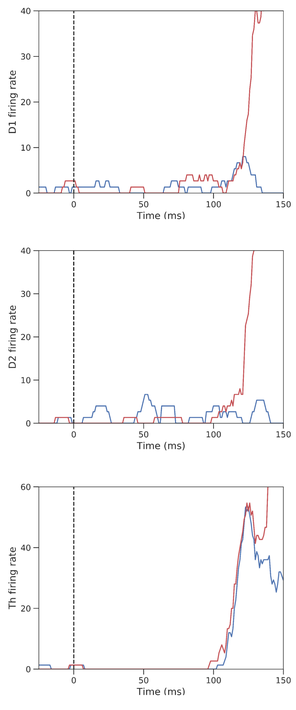
DECISION TIMECOURSE

Start of trial: Equal cortical stimulation applied to both channels

Between decision and reward:
- Cortical stimulation for selected action is maintained
- Unselected action has stimulation removed

Dopaminergic feedback according to Q-value prediction error
D1 and D2 connections undergo spike-timing-dependent plasticity





Key results:
- Striatum and thalamus experience ramping with competition
- Large divergence in D1 and D2 weights in correct directions
- Network behavior influenced by weight balance

Dopaminergic Learning Rule (STDP model)

At each time step [1,2]:

$$\frac{dA_{PRE}}{dt} = \left(-A_{PRE} + \Delta_{PRE} * X_{PRE}(t)\right)/\tau_{PRE},$$
$$\frac{dA_{POST}}{dt} = \left(-A_{POST} + \Delta_{POST} * X_{POST}(t)\right)/\tau_{POST},$$
$$\frac{dE}{dt} = \left(-E - X_{PRE}(t)\, A_{POST} + X_{POST}(t)\, A_{PRE}\right)/\tau_E,$$
$$\frac{dw}{dt} = \alpha_w E f(K_{DA})(w_{max} - w),$$
$$f(K_{DA}) = \begin{cases} K_{DA}, & \text{if the target neuron is a dMSN,} \\ \dfrac{K_{DA}}{c + |K_{DA}|}, & \text{if the target neuron is an iMSN,} \end{cases}$$

If the action is given [3]:

$$DA_{inc}(t) = r_i(t) - \max_i\{Q_i(t)\},$$
$$Q_i(t+1) = Q_i(t) + \alpha\left(r_i(t) - Q_i(t)\right),$$
$$\frac{dK_{DA}}{dt} = \Sigma_i (DA_{inc}(t_i) - K_{DA})\delta(t_i) - K_{DA}/\tau_{DOP},$$
$$\frac{dg_{syn}}{dt} = \Sigma_{i,j} w_i(t_{i,j})\delta(t - t_j) - g_{syn}/\tau_g,$$

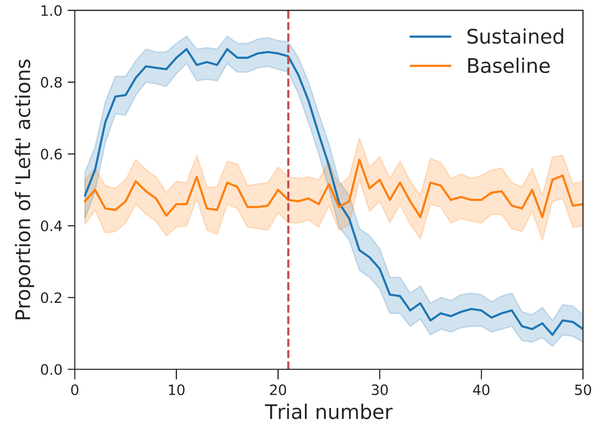
SWITCHPOINT TASK

Switch point task:
- 20 trials in which 'Left' action is rewarded
- 30 trials in which 'Right' action is reward

250 simulations run, time-course of selection probabilities plotted

Key results:
- Successful learning before switch
- Successful relearning of correct action (same asymptotic accuracy)

Lesion study:
- Baseline network without sustained cortical activity shows no learning



SUMMARY & DISCUSSION

Introduction of sustained cortical activity enables effective credit assignment in full cortico-basal-ganglia-thalamic network model, without requiring a change in synaptic-level plasticity mechanisms.

Future directions:
- Explore performance and limitations of model in more complex environments
- Use model to predict physiological activity patterns

REFERENCES & ACKNOWLEDGEMENTS

[1] J. Baladron, A. Nambu, and F. Hamker. *Euro. J. Neurosci.*, doi: 10.1111/ejn.13666, 2017.
[2] K. Gurney, M. Humphries, and P. Redgrave. *PLoS Biol.*, 13(1):e1002034, 2015.
[3] J. Mikhael and R. Bogacz. *PLoS Comp. Biol.*, 12(9):e1005062, 2016.
[4] Yartsev, M., Hanks, T., Yoon, A. & Brody, C. *bioRxiv.*, doi:10.1101/245316, 2018.
[5] Dunovan, K. & Verstynen, T., *Front. Neurosci.* 10.3389/fnins.2016.00106, 2016.
[6] Wiecki, T. V., Sofer, I. & Frank, M. J., *Front. Neuroinform.*, 10.3389/fninf.2013.00014, 2013.
[7] Cisek, P. & J. F Kalaska, *Neuron*, 10.1016/j.neuron.2005.01.02745, 2005.\

Supported by NSF awards DMS 1516288 (CRCNS), 1612913, 1724240 (CRCNS), NSF CAREER award 1351748, and ERDF projects MTM2015-71509-C2-2-R and MTM2017-83568-P.