Dopamine-based reinforcement learning of virtual arm reaching task in a spiking model of cortex

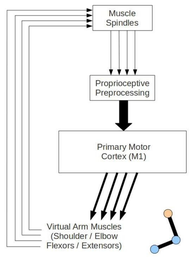
George L. Chadderdon[1], Samuel A. Neymotin[1], Cliff C. Kerr[1,2], Joseph T. Francis[1], William W. Lytton[1,3]

[1] Dept. of Physiology and Pharmacology, SUNY Downstate Medical Center; [2] School of Physics, University of Sydney, Australia; [3] Kings County Hospital, Brooklyn, NY

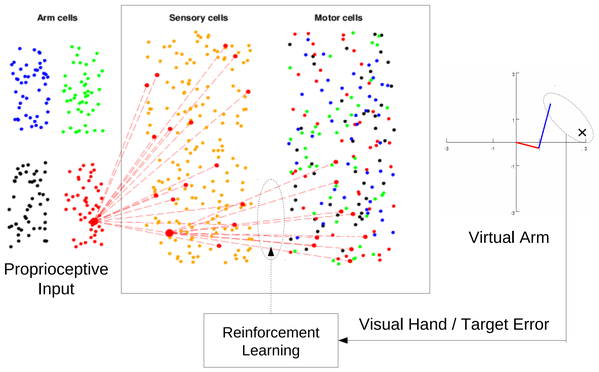
Introduction

Our goal is to model learning and performance in a target-reaching task. We use a spiking model of primary motor cortex to direct a virtual arm toward a target. The model learns by shaping noise-driven "motor babble" into directed motions using a reward / punisher algorithm based on mechanisms from the dopaminergic reward system. The model effectively implements in a spiking model Thorndike's Law of Effect: the proposition that rewards (punishers) make stimulus->response mappings more (less) likely to be triggered in the future.

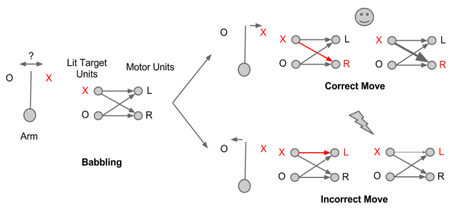
Methods



Continuous loop: Virtual arm provides input via muscle spindles into model which activates arm, which updates spindle activity



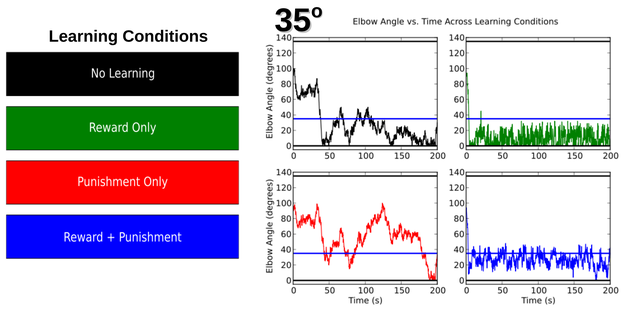
Individual cells in each group project in feedforward direction.



Thorndike Law of Effect: Reward makes behaviors more likely. Punishment makes them less likely.

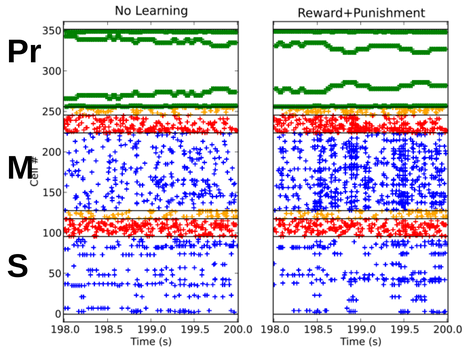
Results

Different types of training



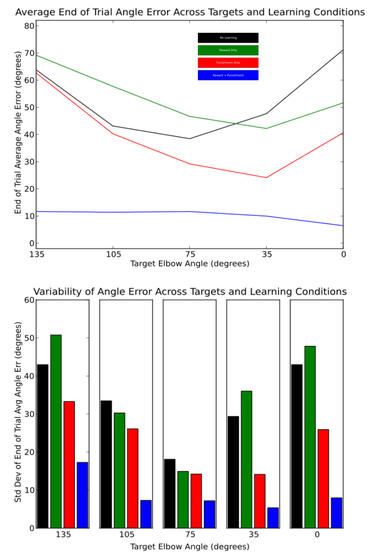
Learning towards target with different conditions: learning is far more rapid with both reward and punishment together

Raster plots



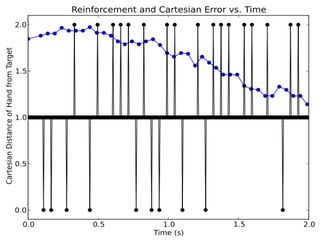
Raster plot at end of learning shows increased firing in the Motor (M) excitatory group (S sensory, Pr proprioceptive)

Consistent results across different wirings (n=500)



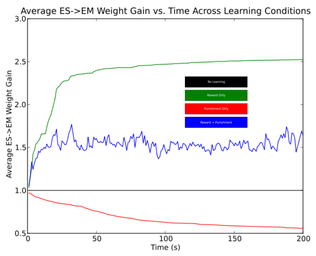
Means and variability of end-of-trial average error are smallest for reward + punishment learning condition

Dopamine signaling



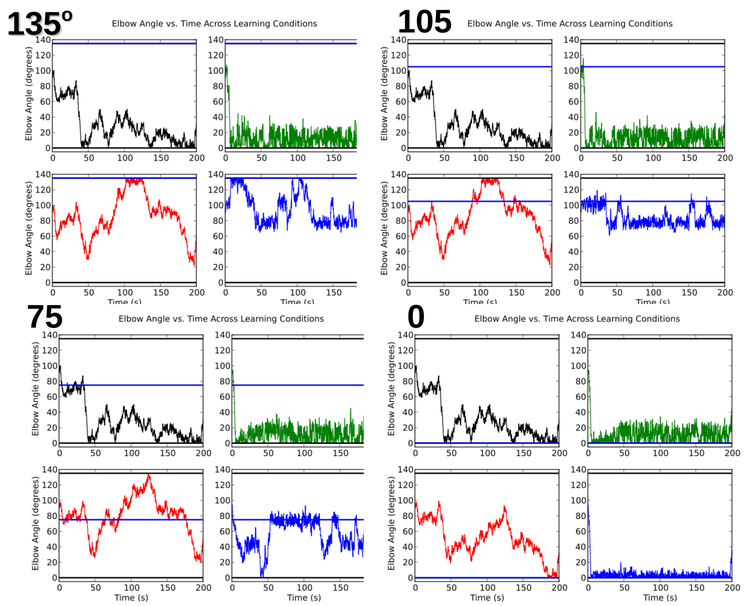
Dopamine signals are given at discrete times allowing error to move towards 0.

Reward / punishment effects on synaptic weights



Weight gains change monotonically unless have both reward and punish

Reach to different targets



Model with both punishment and reward can learn various targets when started at 90 degrees. Note that babbling tends to force hand away from target even after learned.

Conclusions

- **Both reward + punishment are needed for adequate learning**
- **Babble gives tendency to wander noisily from target – plan to improve with adaptive noise mechanism**
- **Plan extending results to 2 degrees of freedom**

References

• Izhikevich EM. 2007. Solving the distal reward signaling. Cereb. Cortex. 17:2443-2452.
• Shen W, Flajolet, M, Greengard P, Surmeier, DJ. 2008. Dichotomous dopaminergic control of striatal synaptic plasticity. Science. 321:848-851.
• Schultz W. 1998. Predictive reward signal of dopamine neuron. J Neurophysiol. 80:1-27.

Acknowledgements

This work was supported by DARPA grant N66001-10-C-2008.