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

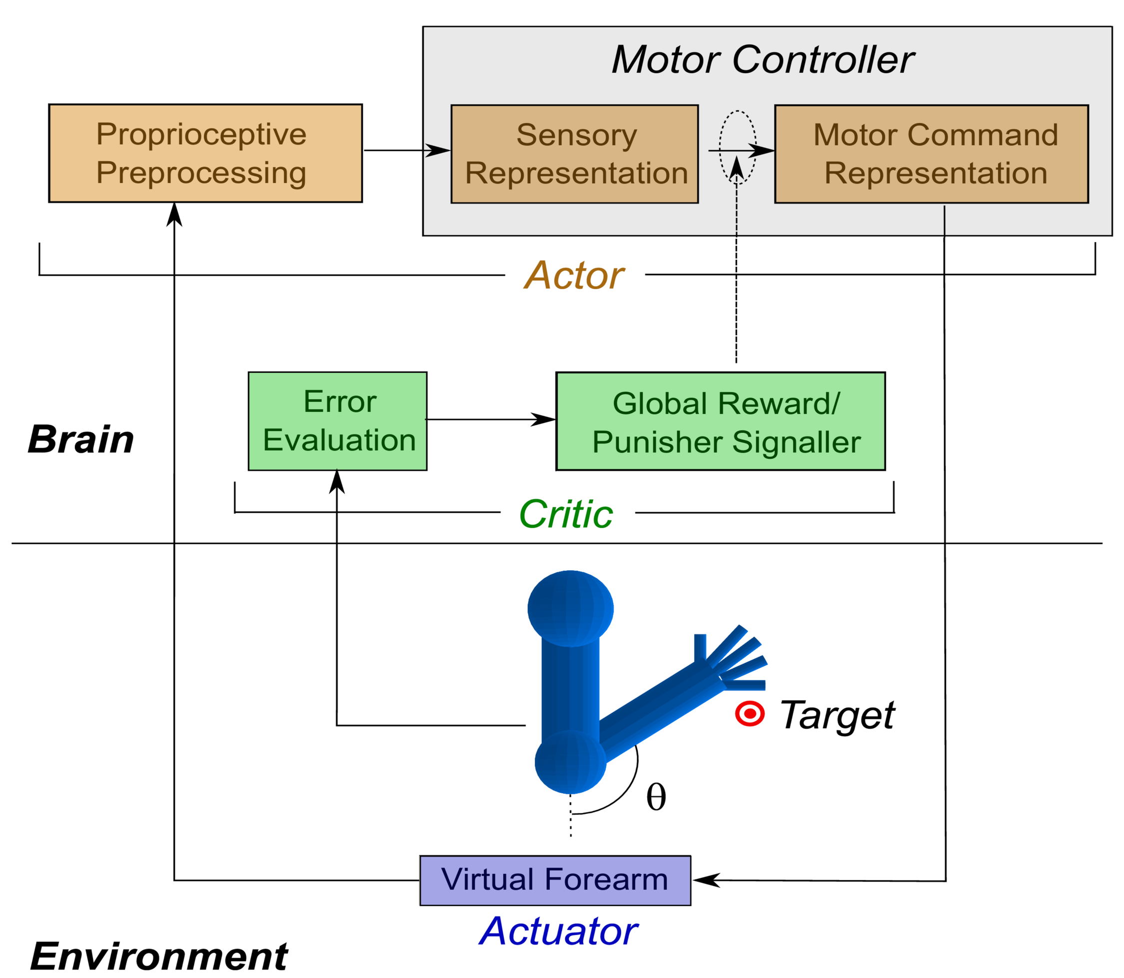
George L. Chadderdon[1], Samuel A. Neymotin[1,2], Cliff C. Kerr[1,3], Joseph T. Francis[1], William W. Lytton[1,4]

[1] Dept. of Physiology and Pharmacology, SUNY Downstate Medical Center; [2] Dept. Of Neurobiology, Yale Univ. School of Medicine; [3] School of Physics, Univ. of Sydney, Australia; [4] Kings County Hospital, Brooklyn, NY

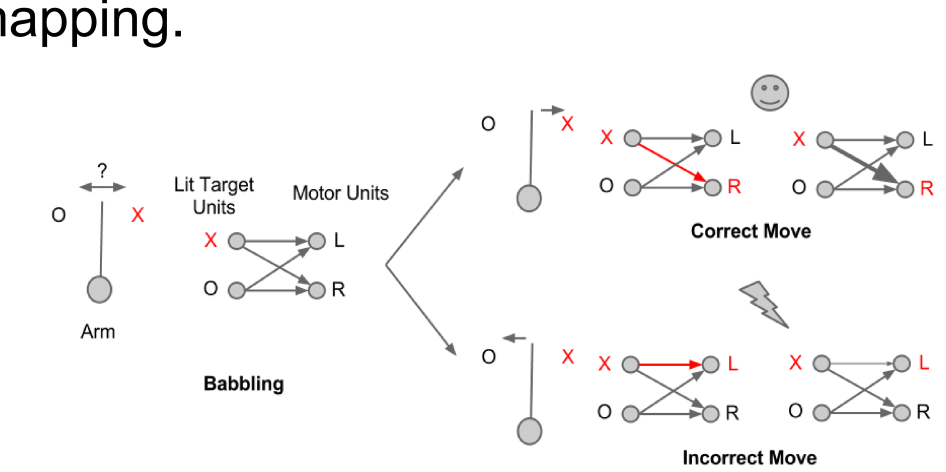
Introduction

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 spiking network model effectively implements Thorndike's Law of Effect: the proposition that rewards (punishers) make stimulus->response mappings more (less) likely to be triggered in the future.

Methods



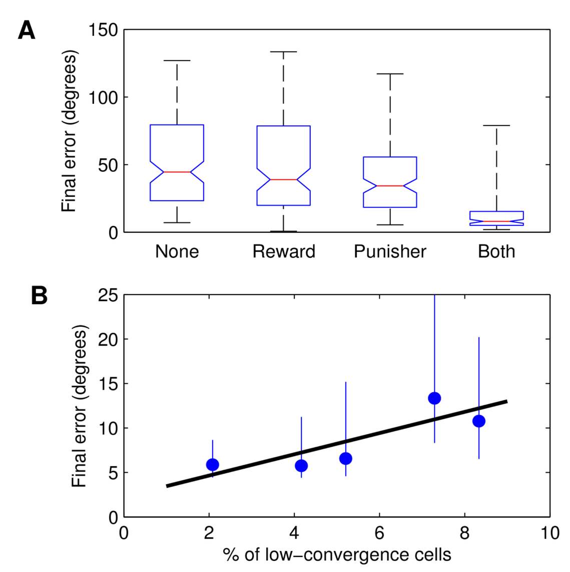
Overview of model and virtual arm system. Only the forearm is allowed to rotate to move the hand toward the target. The arm is driven by a motor controller Actor which is trained by a reward / punisher Critic to learn a proprioceptive sensory->motor command mapping.



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

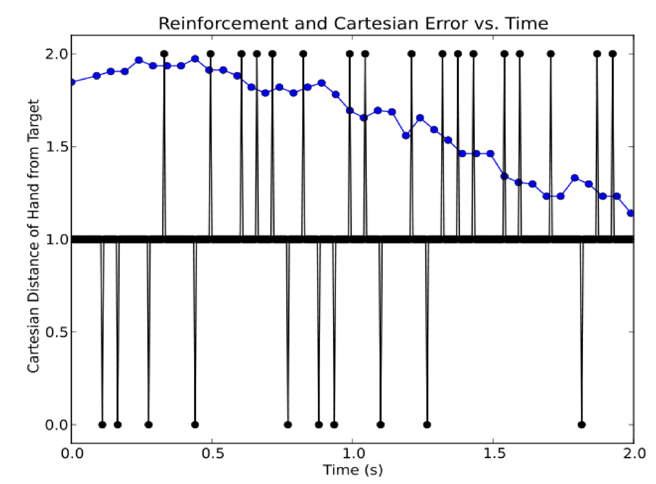
Results

Reaching performance best for reward + punisher



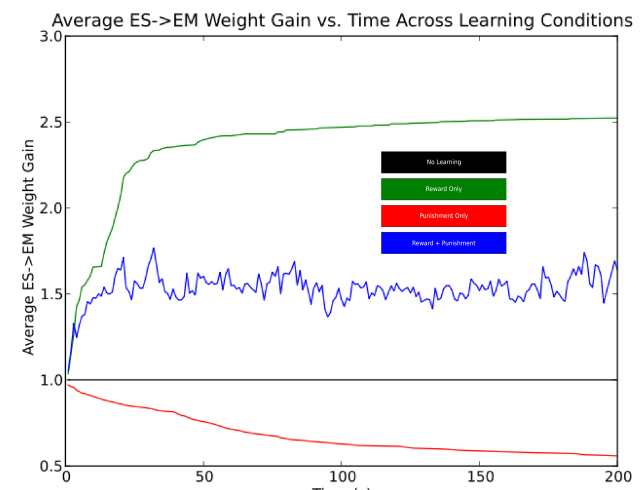
A. Error vs. learning condition. B. Reward + Punisher error vs. % poorly connected neurons.

Dopamine signaling



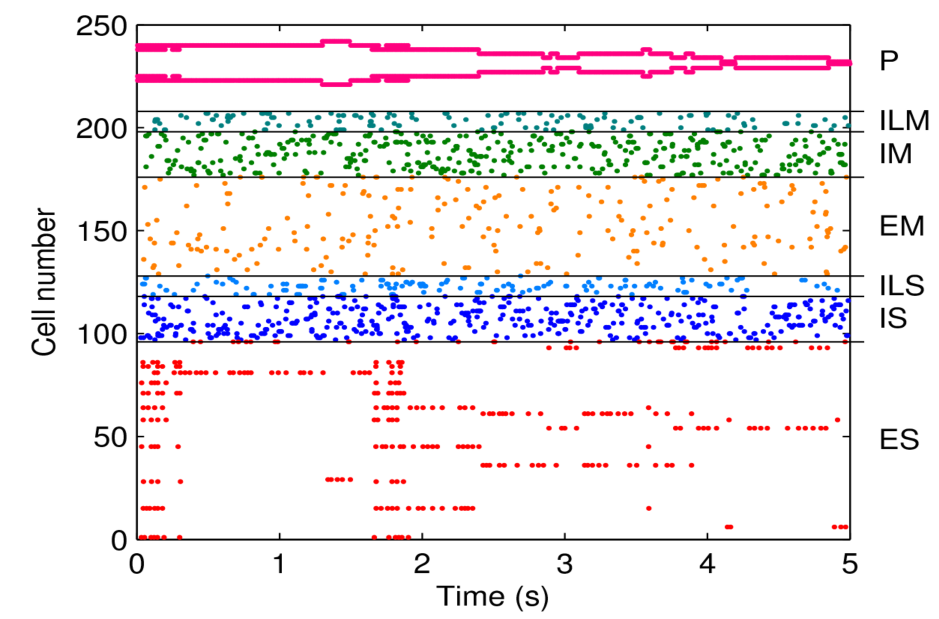
Dopamine signals are given at discrete times allowing error to move towards 0.

Reward / punishment effects on synaptic weights



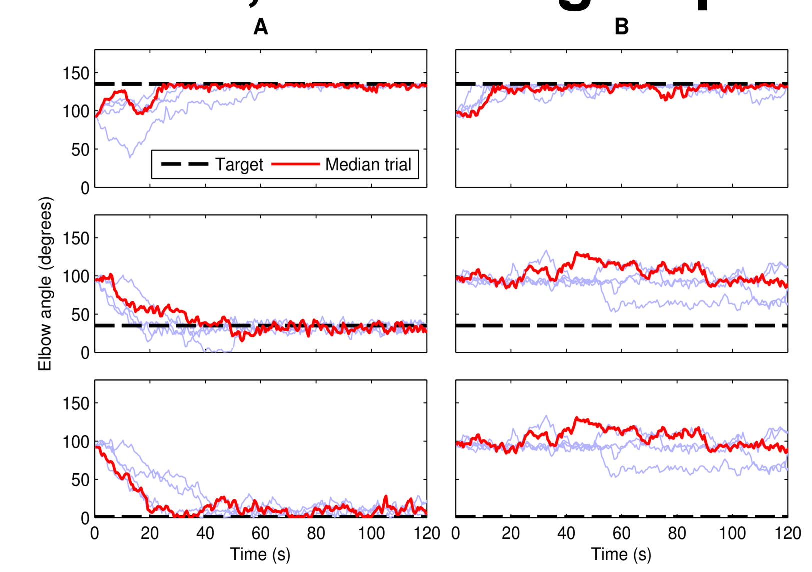
Weight gains change monotonically unless have both reward and punishment.

Raster plot



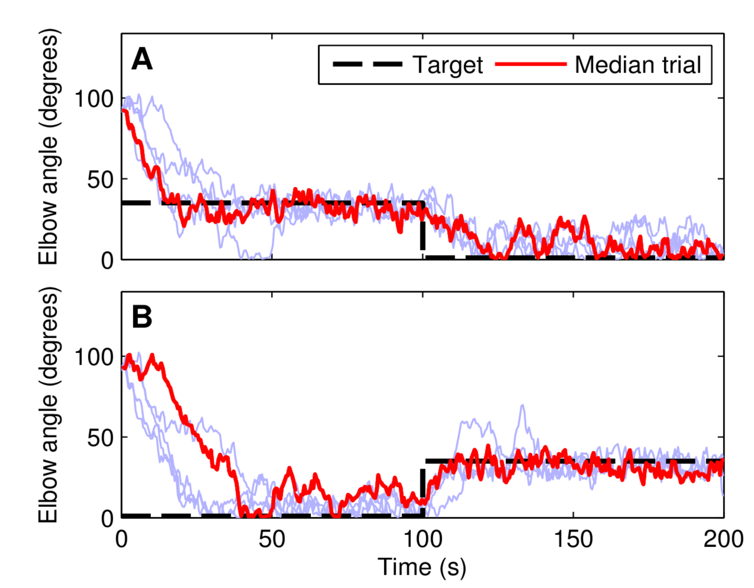
Raster plot during 5 s simulation under no-learning condition.

Reach to different targets successful, but wiring-dependent



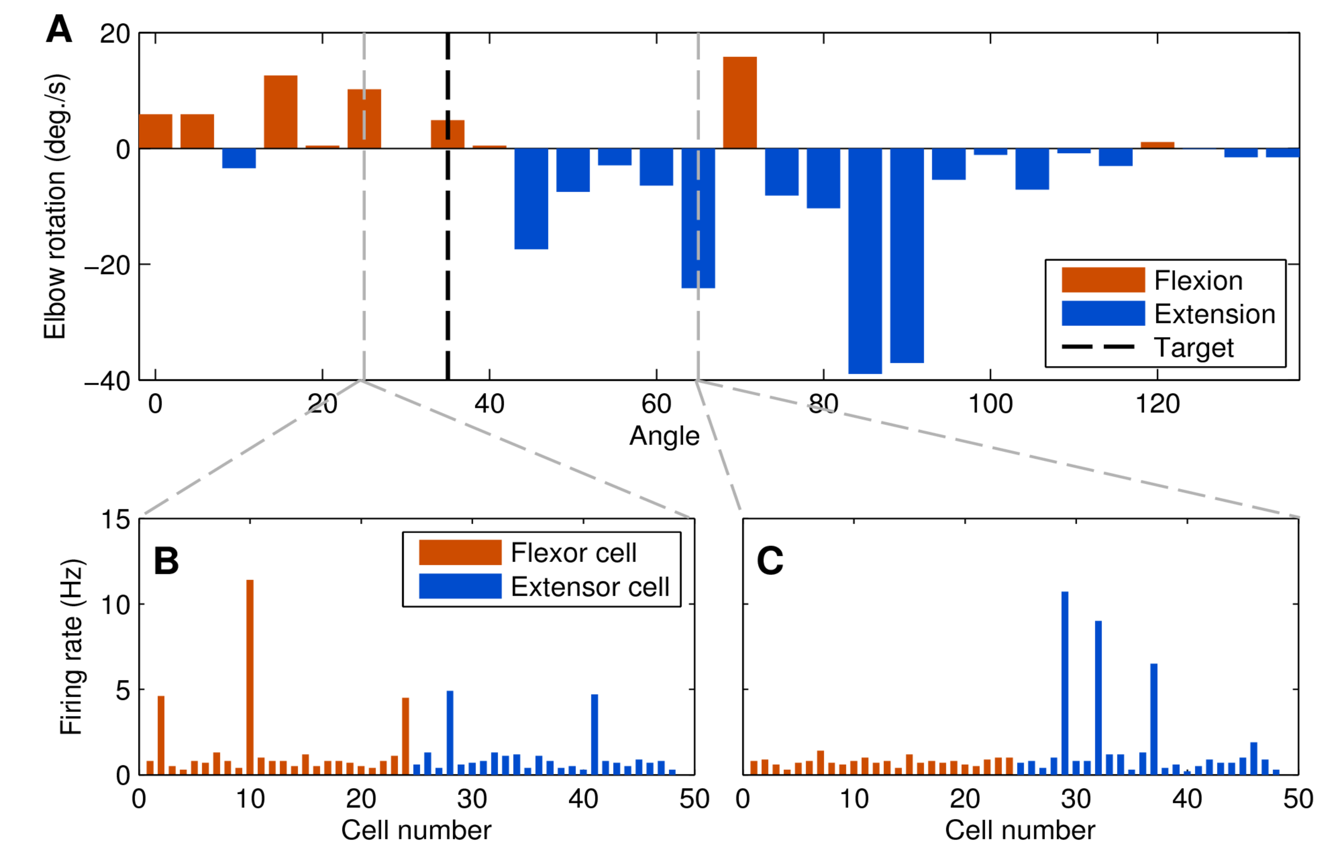
Reach performance on 135 (top), 35 (middle), and 0 (bottom) degrees. A. Good; B. Bad wiring random seed.

Successful target switching performance



A. 35->0 degree switch. B. 0->35 degree switch.

Learning of target attractor at 35°



A. Motor command for trained model vs. arm angle. B. EM cell spiking at 25°. C. EM cell spiking at 65°.

Conclusions

- **Both reward + punishment are needed for adequate learning**
- **Babble allows trial-and-error learning – plan to improve with adaptive noise mechanism**
- **Plan extending model to include cortical laminar structure**

References

• Chadderdon GL, Neymotin SA, Kerr CC, Lytton WW. In press. Reinforcement learning of targeted movement in a spiking neuronal model of motor cortex. PLoS ONE.
• Izhikevich EM. 2007. Solving the distal reward signaling. Cereb. Cortex. 17:2443-2452.
• 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.