

## Learning Reward Timing using Reinforced Consolidation of Synaptic Plasticity.

Harel Z. Shouval<sup>1,3</sup>, Jeff Gavornik<sup>1,2</sup>

<sup>1</sup>Department of Neurobiology and Anatomy the University of Texas Medical School in Houston, TX. <sup>2</sup>Department of Electrical Engineering and <sup>3</sup>Biomedical Engineering the University of Texas. Austin, TX.

Learning interval timing is a crucial component in many behaviors. However, the physiological mechanisms underlying the representation and learning of interval timing have not yet been identified. Recent experimental results indicate that cells within the primary visual cortex can learn to predict the time of rewards associated with visual cues [1]. In this work, different visual cues were paired with rewards at specific temporal offsets. Before training neurons in visual cortex were active only during the duration of the visual cue. However, after sufficient training neurons developed persistent activity beyond the time of the visual cue. The duration of this persistent activity was correlated with the reward time and could be used to predict it.

How can a neural network learn to adapt its temporal dynamics in order to predict an expected reward time? Recurrent connections in a neural network can be tuned in order set different temporal dynamics for neurons within the network [2]. However, it is not clear how a network is able to learn the appropriate recurrent weights. A plasticity model that is able to accomplish this must be sensitive to reward timing, an event that at least initially occurs seconds after the network activity returns to its basal level. Therefore, in order to learn the appropriate dynamics, this network needs to solve a temporal credit assignment problem. In our model plasticity is an ongoing process changing the recurrent synaptic weights as a function of coincident pre- and post-synaptic activity. However, in the absence of reward this plasticity rapidly decays. An external reward allows consolidation of plasticity events that precede a reward, thus reinforcing those specific plasticity events which predict the reward [3]. Additionally we assume that the reward signal is inhibited by the network activity [4]. As a result the network dynamics are altered, acquiring dynamics that are correlated with reward timing. Both abstract and integrate and fire implementations of this network produce dynamics that are similar to experimental results [1].

### Acknowledgments

We thank Marshal Shuler and Mark Bear. This work was supported by an NSF grant: CRCNS - 0515285.

### References

1. Shuler, M.G. and M.F. Bear, *Reward timing in the primary visual cortex*. Science, 2006. 311(5767): p. 1606-9.
2. Seung, H.S., *How the brain keeps the eyes still*. Proc Natl Acad Sci U S A, 1996. 93(23): p. 13339-44.
3. Abbott, L.F. and K.I. Blum, *Functional significance of long-term potentiation for sequence learning and prediction*. Cereb Cortex, 1996. 6(3): p. 406-16.
4. Rescorla, R.A. and A.R. Wagner, *A Theory of Pavlovian conditioning: the effectiveness of reinforcement and non-reinforcement*, in *Classical Conditioning II: Current Research and Theory*, A.H. Black and W.F. Prokasy, Editors. 1972, Appelton-Century-Crofts: New York. p. 64-69.