

A Computational Model of the Role of Orbitofrontal Cortex and Ventral Striatum in Signalling Reward Expectancy in Reinforcement Learning

Robert C. Wilson¹, Yuji K. Takahashi², Matthew R. Roesch⁵, Thomas Stalnaker², Geoffrey Schoenbaum^{2,3} and Yael Niv¹

1. Department of Psychology and Neuroscience Institute, Princeton University, 2. Department of Anatomy & Neurobiology, 3. Department of Psychiatry, University of Maryland School of Medicine, 4. University of Maryland School of Medicine 5. Department Psychology, University of Maryland College Park,

Both the orbitofrontal cortex (OFC) and ventral striatum (VS) have been implicated in signalling reward expectancies. However, exactly what roles these two disparate structures play, and how they are different is very much an open question. Recent results from the Schoenbaum lab (Takahashi et al., this meeting) describing the detailed effect of OFC lesions on putative reward prediction error signalling by midbrain dopaminergic neurons of rats, point to one possible delineation.

Here we describe a reinforcement learning (RL) model of the Takahashi et al. results, that suggests related, but slightly different roles for the OFC and VS in signalling reward expectancies. We present an actor/critic model with one actor (putatively the dorsal striatum) and two critics (OFC and VS). We hypothesise that the VS critic learns state values relatively slowly and in a model free way, while OFC learns state values faster and in a model based way, using one step look ahead. Both areas contribute to a single prediction error signal, computed in ventral tegmental area (VTA), that is used to teach both critics and the actor. As they receive the same teaching signal, the critics, OFC and VS, essentially compete for the value of each state.

Our model makes a number of predictions regarding the effects of OFC and VS lesions on the response properties of dopaminergic (putatively prediction error encoding) neurons in VTA. The model predicts that lesions to either VS or OFC result in persistent prediction errors to predictable rewards and diminished prediction errors on the omission of predictable rewards. At the time of a reward predicting cue, the model predicts that these lesions cause both positive and negative prediction errors to be diminished.

When the animal is free to choose between a high and low valued option, we predict a difference between the effects of OFC and VS lesions. In the “unlesioned” model, because of the proposed look-ahead abilities of OFC, the model predicts differential signals at the time at which the decision is made, corresponding to whether the high or low valued option has been chosen. When the model-OFC is “lesioned”, however, these differential signals disappear as the model is no longer aware of the decision that will be made. This is not the case when model-VS is “lesioned” in which case the difference between high and low valued options persists. These predictions regarding OFC lesions are born out in Takahashi et al.’s experiments on rats.