Neurons of the prefrontal cortex encode a representation of a Bayesian belief during reinforcement learning

Ramon Bartolo (ramon.bartoloorozco@nih.gov)
Laboratory of Neuropsychology, NIMH/NIH, 49 Convent Drive, Room 1B80
Bethesda, MD 20892-4415 USA

Richard C. Saunders (richardsaunders@mail.nih.gov)
Laboratory of Neuropsychology, NIMH/NIH, 49 Convent Drive, Room 1B80
Bethesda, MD 20892 USA

Andrew R. Mitz (mitza@mail.nih.gov)
Laboratory of Neuropsychology, NIMH/NIH, 49 Convent Drive, Room 1B80
Bethesda, MD 20892 USA

Bruno B. Averbeck (averbeckbb@mail.nih.gov)
Laboratory of Neuropsychology, NIMH/NIH, 49 Convent Drive, Room 1B80
Bethesda, MD 20892 USA

Abstract:

In a noisy environment, learning requires organisms to keep track of choices and their associated outcomes across successive decisions to form beliefs about value in the world. This form of reinforcement learning allows them to predict future outcomes and to update their belief after each outcome. The primate prefrontal cortex (PFC) integrates information carried by reward circuits, in addition to its role in working memory. Also, it has been suggested that the prefrontal cortex plays a critical role in inferring the current state of the world under some level of uncertainty. In the present study, we explore the PFC computations related to updating current beliefs in multifaceted reward environments. We conducted high-channel count single-unit recordings in two male macaques while they executed a two-armed bandit reversal learning task. Behavioral analyses showed that they used this prior knowledge to guide their choice preference. We found activity associated to posterior probability estimates. Overall these results suggest that prefrontal neurons encode decisions associated with Bayesian subjective values and highlight the role of the PFC in representing a belief about the current state of the world.

Keywords: PFC; Reinforcement Learning; Neurophysiology.

Introduction

In a noisy environment, learning requires organisms to keep track of choices and their associated outcomes across successive decisions to form beliefs about value in the world. This form of reinforcement learning allows them to predict future outcomes and to update their belief after each outcome. The primate prefrontal cortex (PFC) integrates information carried by reward circuits, in addition to its role in working memory. Also, it has been suggested that the prefrontal cortex plays a critical role in inferring the current state of the world under some level of uncertainty. In the present study, we explore the Prefrontal Cortex (PFC) computations related to the acquisition and update of beliefs in multifaceted and dynamical reward environments.

Experimental setup

We trained two adult macaques in a two-armed bandit reversal learning task. While the monkeys were performing the task the activity of neurons from the prefrontal cortex was recorded using a bilateral implant of 8 Utah arrays (10X10 layout, 96 electrodes each), 4 on each hemisphere. Then we combined behavioral analysis using a Bayesian framework with the analysis of the neural activity.

Two-armed bandit, reversal learning task.

The task had two different types of trials (WHAT and WHERE) and was organized in blocks of 80 trials of the same type. On each trial, the monkey was required to hold central fixation for a variable time (400-600ms). After the fixation period, two images were presented to the left and right of the fixation spot and the animal had to choose one by making a saccade towards the chosen image. The location of each image was randomized between trials. On WHAT trials, each image was associated with a fixed probability for reward delivery, and one image had a higher reward probability, making it the best option to maximize reward. On WHERE trials reward contingencies were determined only by the direction of the saccade, independently of the image chosen. These types of trials are schematized in Figure 1. On each block of trials, a new set of images was used. Though trial-and-error, the animals had to discover what type of trials composed the block (Block Type) and which of the two targets (images or locations) had the highest reward probability. Reward mappings across targets were swapped across targets at a randomly determined trial within an interval (trials 30-50) and the animals had to reverse their choices to maximize reward. The animals were extensively trained in this task before the electrophysiological recordings, hence, this reversal was expected but the animals could not predict exactly when it had occurred.

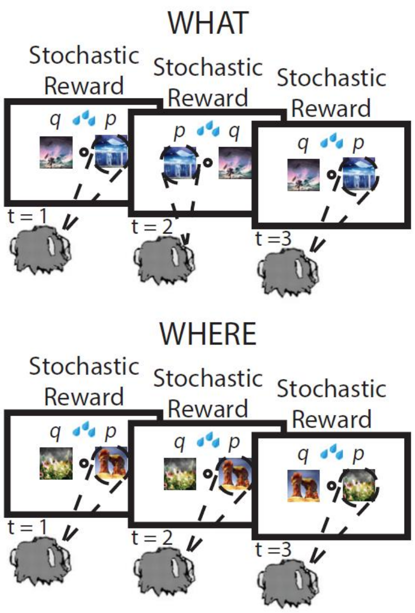

Figure 1: Schematic of the two types of trials in the reversal learning task.

Bayesian analysis.

We used a previously described Bayesian model (Costa et al, 2015; Rothenhoefer et al, 2017) to dissect the behavior and extract the distribution of posterior probabilities across trials for Block Type. The model was also used to estimate the distribution of the posterior probability that reversal had occurred.

Analysis of the neural activity.

We aligned the neural activity to the onset of the targets on each trial of the reversal learning task. Then, we counted spikes within a sliding window (300ms width, 25ms step) that moved around the target onset time (-1500ms to 1500ms) to get a series of 109 bins for each trial and neuron. Then we constructed an ANOVA model to assess the association between the neural activity and a series of behavioral parameters. We included 7 behavioral parameters as regressors in the model: 1) location chosen on each trial (left or right), 2) image chosen (A or B), 3) trial outcome (reward or no reward), 4) block number within the recording session, 5) trial number within block, 6) cumulative posterior probability for Block Type and 7) posterior probability that a reversal had occurred. We applied this ANOVA model to each neuron and spikes bin across all trials. We report the fraction of cells with a significant main effect of each factor across time around target onset.

Results

Results are summarized in Figure 2. We found a strong association between the neural activity and both the location chosen and the image chosen. We also found that the neural activity is associated with the block number within a session, even before the presentation of the targets, suggesting that the PFC neurons maintain a representation of the block currently being executed. In addition, we also found a strong association with the outcome of the trial towards the end of the trial. Remarkably, we also found activity associated with the Bayesian estimates for Block Type and Reversal Probability. Overall these results suggest that prefrontal neurons encode oculomotor decisions associated with Bayesian subjective values and highlight the role of the PFC in representing a belief about the current state of the world.

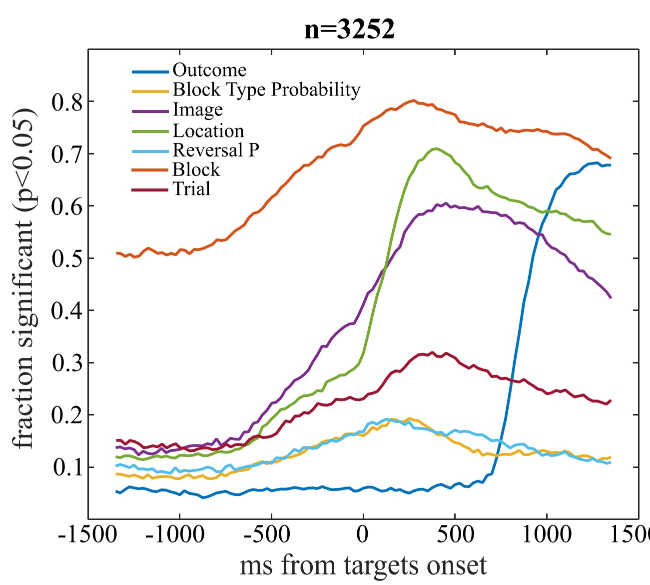

Figure 2: Fraction of neurons with significant main effects of the factors listed in the legend.

References

Costa, V.D., Tran, V.L, Turchi, J., & Averbeck, B.B. (2015). Reversal Learning and Dopamine: A Bayesian Perspective. *The journal of Neuroscience*, 35(6), 2407-2416.

Rothenhoefer, K.M., Costa, V.D, Bartolo, R., Vicario-Feliciano, R., Murray, E.A., & Averbeck, B.B. (2017). Effects of Ventral Striatum Lesions on Stimulus-Based versus Action-Based Reinforcement Learning. *The journal of Neuroscience*, 37(29), 6902-6914.