Neural dynamics of reversal learning in the prefrontal cortex

Christopher M. Kim, Carson C. Chow
Laboratory of Biological Modeling, NIDDK/NIH, Bethesda, MD, USA

Bruno B. Averbeck
Laboratory of Neuropsychology, NIMH/NIH, Bethesda, MD, USA

Abstract

We used a reversal learning task with probabilistic reward to investigate how a neural network accumulates evidence across multiple trials to reverse its decision. We analyzed prefrontal cortex activity in monkeys performing the task and recurrent neural networks trained to learn the behavioral strategies of monkeys. We found substantial neural dynamics across the time span of a trial in the subspace that encodes reversal probability. This suggested that the standard attractor model for evidence accumulation, in which network states do not deviate strongly from attractor states, does not explain the observed neural activity. We found that reward outcomes affected the entire reversal probability trajectories systematically. The reversal probability trajectories across trials had temporally stable ordering, and the reversal trial was decodable over a wide time span. These findings show that, when performing a task that requires intervening behavior, reversal probability activity across trials is encoded in dynamic neural trajectories, allowing temporally flexible representation of decision-related evidence.

Introduction

Reversal learning has been used for assessing the ability to adapt one's behavior in a dynamically changing environment (Costa, Tran, Turchi, & Averbeck, 2015; Groman et al., 2019; Bartolo & Averbeck, 2020). In these tasks, there is uncertainty in when to reverse one's choice, as reward is received stochastically even when the less favorable option is chosen. Therefore, it is essential that reward outcomes are integrated over multiple trials before the decision reversal.

In this study, we investigated the neural dynamics that underlie multi-trial evidence accumulation in the reversal learning task. We found that intervening behavior during a trial produced substantial non-stationary neural activity. This made the attractor dynamics (Wong & Wang, 2006; Mante, Sussillo, Shenoy, & Newsome, 2013; Luo et al., 2023), which is a standard neural model for evidence accumulation and require the network state to remain close to attractor states, ill-suited for explaining the neural activity associated with evidence accumulation in reversal learning.

Here, we developed a neural network model that learns the behavioral strategies of monkeys (Costa et al., 2015; Bartolo & Averbeck, 2020) and found that a family of graded neural trajectories that evolve dynamically in time encodes the reversal probability in the trained network and prefrontal cortex. These findings suggest that, in tasks that require executing intervening behavior, evidence across multiple trials is accumulated in the form of dynamic trajectories that allow for temporally flexible representation.

Methods

Reversal learning task

Each block consisted of $T = 24$ trials during network training. The reversal trial r was sampled randomly and uniformly from 10 trials around the midtrial $r \in Unif[T_m - 5, T_m + 5], T_m = T/2$. To model the reversal of stochastic rewards, two targets were generated at each trial k with probability $Pr(trg = 0) = q_k$ and $Pr(trg = 1) = 1 - q_k$ with $q_k = p$ before the reversal ($k < r$) and $q_k = 1 - p$ after the reversal ($k \geq r$). Network's choice was compared to the target, and one of four types of feedback inputs, based on the (choice, matched)-pair, was provided to the network: (0, not matched), (0, matched), (1, not matched) and (1, matched) (**Fig.1A**).

Bayesian inference model

Ideal observer model The ideal observer model infers the experimentally scheduled reversal trial and assumes that (a) target probability is known, and (b) it switches at a scheduled reversal trial.

The data available to the ideal observer are the choice $y_k \in \{0, 1\}$ and the reward outcome $z_k \in \{0, 1\}$ at all the trials $k \in [1, T]$. Then, the posterior distribution of scheduled reversal at trials $k \in [1, T]$ can be inferred from Bayes' rule $p(r|y,z) = p(y,z|r)p(r)/p(y,z)$, where the likelihood function $f_{IO}(r) = p(y_{1:t}, z_{1:t}|r)$ is defined by $f_{IO}(r) = \prod_{k=1}^{t} q_k$. Here, $q_k = p$ if $(y_k, z_k) = (0, 1)$; $q_k = 1 - p$ if $(0, 0)$; $q_k = 1 - p$ if $(1, 1)$; $q_k = 1 - p$ if $(1, 0)$ before the reversal, i.e., $k < r$. After the reversal, i.e., $k \geq r$, the reward schedule is switched.

Behavioral model To infer the trial at which the preferred choice switches, i.e., behavior reversal, we applied the same framework as the ideal observer, but used a likelihood function that assumes a switch only in the preferred choice probability but not in the reward schedule.

Recurrent neural networks

Initial network The initial network was a recurrent neural network with purely inhibitory synaptic connections with a baseline external input to sustain the network activity. Such inhibitory network operated in a balanced regime where the recurrent inhibitory inputs were balanced with the external excitatory inputs (van Vreeswijk & Sompolinsky, 1996).

Training scheme To learn the behavioral strategies of monkey, we trained the network to learn from outputs of the ideal observer. The network was first simulated and then its choices and reward outcomes were fed into the ideal observer to infer the scheduled reversal trial. Then, the network was trained to switch its preferred choice a few trials after the inferred reversal trial. This delay in the behavior reversal was observed in monkey reversal behavior (Bartolo & Averbeck, 2020; Costa et al., 2015). As the scheduled reversal trial varied across blocks, the network learned to reverse its choice in a block-dependent manner.

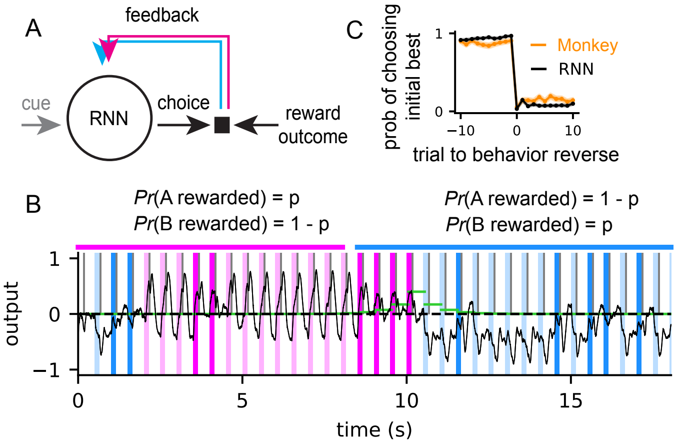

Figure 1: Recurrent neural network trained on ideal observer

Results

Here we show that RNNs trained on the ideal observer show behavior outputs (**Fig.1**) and neural activity (**Fig.2**) similar to monkeys performing the same task. We found that substantial temporal dynamics are produced in the neural subspace that encode evidence, i.e., reversal probability, and their changes are driven by reward outcomes (**Fig.2**). Also, perturbation experiments show that the reversal probability activity may systematically change choice activity, suggesting causal link between the two variables.

Trained network behavior

After training, RNNs were resistant to a few no-reward feedback trials but abruptly switched its choice (magenta to blue) when consecutive no-reward feedback was received (**Fig.1B**). Dark (light) magenta shows no-reward (reward) trials when option A is chosen. Similarly, dark (light) blue shows reward outcomes when option B is chosen. Such abrupt switch in choice was consistent with monkey's behavior (**Fig.1C**).

Reward shapes the reversal probability activity

Next, we analyzed the neural activity of PFC neurons and trained RNNs. In particular, we performed targeted dimensionality reduction (Mante et al., 2013) to identify neural subspace that encode choice and reversal probability activity. We found that substantial temporal dynamics were observed in the two-dimensional subspace $(x_{choice}, x_{reverse})$, and the neural trajectories shifted systematically across trials (**Fig2A**).

We asked if changes in reversal probability activity was driven by reward outcomes. To investigate, we set up an integration equation $x_{reverse}(k+1) = x_{reverse}(k) + R^{\pm}(k)$, where $R^{\pm}(k)$ is estimated based on reward $(+)$ or no-reward $(-)$ in trial k, and were able to predict the reversal probability activity at cue offset in upcoming trials (**Fig.2B**, right).

Given the substantial neural dynamics in this subspace over time within a trial, we inquired if the reward outcomes could shape the entire neural trajectory. We found that, when reward was received, the $x_{reverse}(t)$ trajectory was shifted upward across the time span of a trial. On the other hand, no-reward led to a downward shift in the $x_{reverse}$ trajectory (**Fig2C**, left). Moreover, consecutive no-reward (reward) outcomes increased (decreased) the ramping rate of $x_{reverse}$ towards cue offset (**Fig2C**, right).

These findings suggest that evidence for reversing a decision, i.e., reversal probability, is not encoded in static stationary states, but in dynamic trajectories that evolve in time and shift across trials. Furthermore, we analyzed the ordering of $x_{reverse}$ trajectories across trials at each time point and found their ordering is stable in time, suggesting temporally stable representation of the dynamic trajectories (**Fig.2D**).

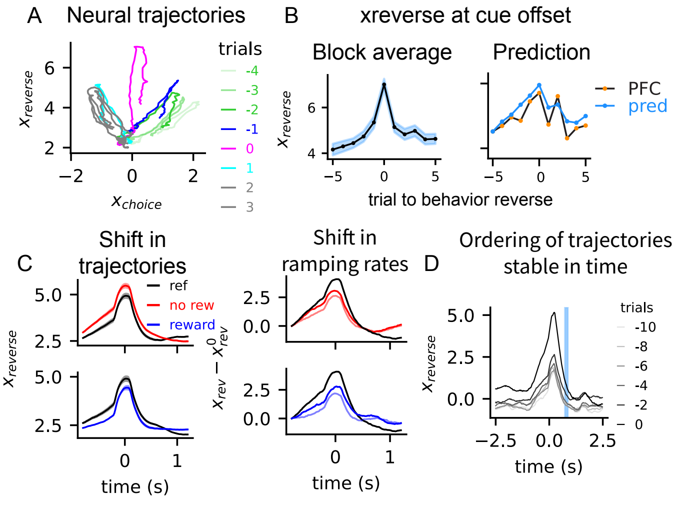

Figure 2: Neural activity of prefrontal cortex of monkey

Perturbing reversal probability activity

To test if reversal probability activity is causally linked to choice outcomes, we perturbed the neural activity of trained RNNs along the direction encoding reversal probability (v_+), against it (v_-) and randomly (v_c). We found that, when reversal probability activity is decreased (increased), the reversal trial was delayed (accelerated), suggesting perturbing reversal probability activity can systematically affect the choice outcomes. Although, perturbation experiment was not performed in monkey PFC, we analyzed the residual activity of $x_{reverse}$ and

x_{choice} around their block-averaged activity, consdering them as “natural” perturbation, and found that they were negatively correlated, consistently with RNN results.

Acknowledgments

This work was supported by the Intramural Program of the NIDDK and NIMH at NIH.

References

Bartolo, R., & Averbeck, B. B. (2020). Prefrontal Cortex Predicts State Switches during Reversal Learning. *Neuron*, *106*(6), 1044–1054.e4. doi: 10.1016/j.neuron.2020.03.024

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal learning and dopamine: a bayesian perspective. *Journal of Neuroscience*, *35*(6), 2407–2416.

Groman, S. M., Keistler, C., Keip, A. J., Hammarlund, E., DiLeone, R. J., Pittenger, C., ... Taylor, J. R. (2019). Orbitofrontal circuits control multiple reinforcement-learning processes. *Neuron*, *103*(4), 734–746.

Luo, T. Z., Kim, T. D., Gupta, D., Bondy, A. G., Kopec, C. D., Elliot, V. A., ... Brody, C. D. (2023). Transitions in dynamical regime and neural mode underlie perceptual decision-making. *bioRxiv*, 2023–10.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, *503*(7474), 78–84.

van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, *274*(5293), 1724.

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, *26*(4), 1314–1328.