The causal contributions of medial prefrontal cortex to value-based decisions in mice

Huriye Atilgan (huriye.atilgan@yale.edu)
Department of Psychiatry, Yale School of Medicine
New Haven, CT, USA

Cayla Murphy (cayla.murphy@yale.edu)
Department of Psychiatry, Yale School of Medicine
New Haven, CT, USA

Alex C. Kwan (alex.kwan@yale.edu)
Department of Psychiatry, Yale School of Medicine
New Haven, CT, USA

Abstract:

Learning from experience is essential to the optimization of behavior. In particular, we learn from past choices and outcomes to infer the predicted values of the actions to be taken. Then based on the values, we may select an informed choice. However, despite the many neural correlates identified, we still do not have a clear picture for how values are computed and translated into informed behavior. Here, we trained head-fixed mice to perform a two-armed bandit task. Animals based their decisions on past choices and reinforcements, consistent with having an internal representation of action values. To determine the causal contributions of the medial prefrontal cortex, we tested the animals before and after an excitotoxic lesion of the medial secondary motor cortex (M2). We found that unilateral M2 lesion led to side-specific effects on the animal's ability to learn from past choices. To quantify the decision-making process, we fitted the animal's choice behavior with Q-learning models to extract learning parameters such as learning rate, forgetting rate, and inverse temperature. Altogether, the results provide insights into the causal involvement of mouse mM2 in value-based decision making.

Keywords: value-based decision making; medial prefrontal cortex; reinforcement learning

Introduction

The assignment of values to different actions and situations is a core component of reinforcement learning and decision-making. Therefore, it is not surprising that a distributed circuitry involving many brain regions have been implicated in the calculation, representation, and use of values (Corrado and Doya, 2007; Lee et al., 2012). The medial prefrontal cortex (mPFC) is thought to a final common path for integrating and comparing value signals for action selection (Levy and Glimcher, 2012).

Bilateral lesion or inactivation of rodent medial secondary motor cortex (M2) impairs the learning and flexible use of reward-guided choices (Makino, 2017; Siniscalchi, 2016; Sul, 2011), which are essential to update estimated values of action.

However, the previous studies had two shortcomings. One, M2 was silenced for both hemispheres. Deficits due to inability to learn from rewarded and/or unrewarded choices cannot be easily dissociated. Two, in an environment with changing volatility, reward probabilities may be stable or fast-changing, and it is known that an agent may adjust the learning rate to match the volatility of the environment. The contribution of M2 to learning rate adjustment is unclear. Here, using unilateral excitotoxic lesion, we characterized the causal contribution of M2 on learning from past choices in an environment with changing volatility.

Results

Animals based their decisions on past choices and reinforcements

Head-fixed mice were trained on a dynamic foraging task based on a two-armed bandit design. For each trial, a mouse would wait for an auditory cue and then make a left or right choice by tongue lick. The outcome is based on the reward probabilities (e.g. "10:70": 10% chance to receive water for a left choice; 70% for right, Figure 1A). Animals would perform a number of trials, until the reward probabilities switch (e.g., "70:10"). The switching criterion is 10 trials with choices on the high-probability side plus a random number. The reward probabilities would continue to switch until the end of the session. Animals thus have to continually learn from its past choices and outcomes to maximize the number of rewards in the task.

In the example session involving two sets of reward probabilities, animal performed more than 400 trials, including 17 reversals (**Figure 1**). Animals chose the high-probability side more frequently and, following a block switch, would quickly switch their preferred action. Actions followed by rewards were more likely to be subsequently selected, as indicated by a logistic regression analysis (data not shown).

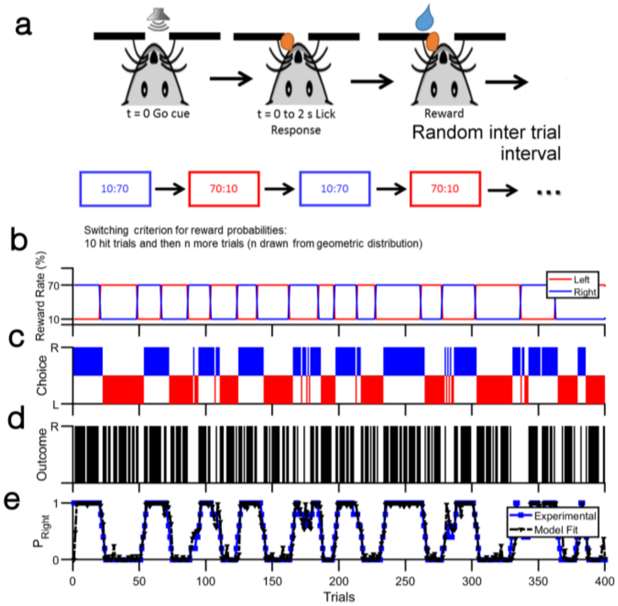

Figure 1: Schematic representation of task design and performance of a mouse in one behavioral session

Because switches occur after varying numbers of trials, we could characterize the animal's choice behavior in situations with different degrees of volatility. Animals reversed slower after a long block (> 40 trials before block switch), relative a short block (<25 trials, **Figure 2a-b**), suggesting that they are sensitive to the volatility of the environment and adjust their current learning rate accordingly.

In other sessions, animals were tested on six different sets of reward probabilities including 70:10, 70:30, 30:10, 30:70, 10:30, and 10:70. Animals were quicker to switch when the reward probability difference is greater (i.e., quicker switch from 70:10 to 10:70 compare to 10:30; **Figure 2c**), indicating that they based their decisions on past choices and reinforcements, consistent with having an internal representation of action values.

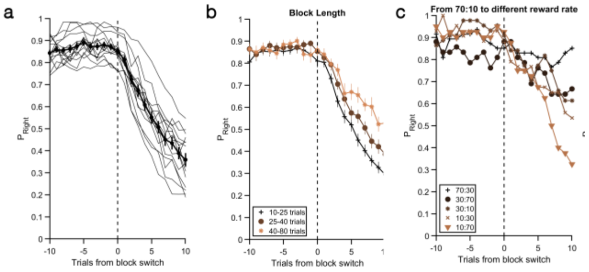

Figure 2: The mean ± s.e.m) probability for choosing the different sides for trials around a block switch (n=20 mice) for different block lengths and different reward rates.

After M2 lesion, animals were quicker to switch contralateral to lesion side

Ibotenic acid lesion was performed after behavioral testing for at least 7 sessions. After two weeks of recovery, the animals (n=6) were tested again. Animals with unilateral lesions were slower in switching to the ipsilateral side compared to their pre-lesion data (**Figure 3a-b**). By contrast, they were quicker in abandoning the ipsilateral side when the associated reward probability was reduced. Additionally, block length influence on their switch was reduced, suggesting diminished sensitivity to volatility (**Figure 3c**). These results suggest either reduced learning from rewarded contralateral actions or facilitated learning from unrewarded ipsilateral actions.

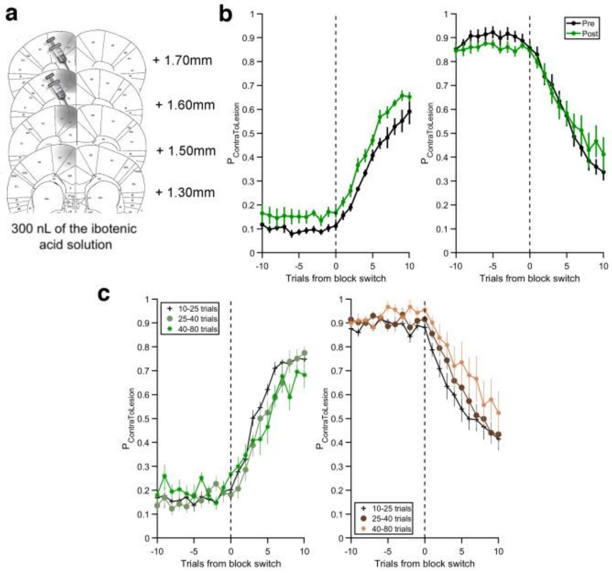

Figure 3: Lesioned area and lesion results (n=6)

Discussion and Future Work

The results showed that a unilateral M2 lesion led to side-specific effects on the animal's ability to learn from past choices.

We fitted the pre-lesion data to different Q-learning algorithms and found that a DFQ-learning model in which the learning rate for unrewarded actions is the same as the forgetting rate provided the best fit (**Figure 4**) and captured the choice behavior (**Figure 1e**). This procedure will allow us to quantitatively extract the decision-related parameters including decision value, chosen values, action values, reward prediction error, learning rate, and inverse temperature. By comparing with the decision-related parameters extracted from the lesion data, we may gain further insights into the causal contribution of medial M2 on learning from past choices in a volatile environment.

Methods

Two armed bandit task

Six adult male mice with a C57Bl/6J genetic background were used in the presented data. The mice were placed in a modified acrylic tube and held head-fixed during the bandit task by fastening their surgically implanted head plate to a stainless-steel bracket. In order for the mice to be motivated to perform the task, subjects were water restricted. Water was only provided in the form of rewards in a single daily training session.

Before each trial, the mouse was required to suppress licking for a random duration. Trial would begin with an auditory cue (5 kHz, 500 ms). The mouse could indicate its response with a tongue lick to the left or right spout. Depending on the reward probabilities, the outcome might be 2 μL of water. At 3 s after the outcome, a no-lick period began until the next trial. Upon licking the high-probability side for ten trials plus a randomly selected number of trials, the reward probabilities would switch.

Excitotoxic lesion

A volume of 300 nL of the ibotenic acid (505024; Abcam) solution was injected into M2 (left or right medial secondary motor region: AP: +1.5 mm; ML: ±0.3 mm; DV: 0.3 mm, **Figure 3A**). Three mice received the injections on the M2 region in the left hemisphere and three on M2 region in the right hemisphere.

Reinforcement Learning Model

To model the learning process, we used Q-learning algorithms (Ito and Doya, 2009). The action value Q_i (t), which is the value for an action i ∈ $\{L, R\}$, is updated by the following:

$$Q_i(t+1) =$$

$$\begin{array}{ll} Q_i(t) + \alpha_{Ri}\ \big(r(t) - Q_i(t)\big) & \text{i}f\ \text{a}(t) = \text{i},\ r(t)\text{=1} \\ Q_i(t) + \alpha_{URi}\big(r(t) - Q_i(t)\big) & \text{i}f\ \text{a}(t) = \text{i},\ r(t)\text{=0} \\ (1-\lambda_1)\ Q_i(t) & \text{i}f\ \text{a}(t) \neq \text{i},\ r(t)\text{=1} \\ (1-\lambda_2)\ Q_i(t) & \text{i}f\ \text{a}(t) \neq \text{i},\ r(t)\text{=0} \end{array}$$

where a(t) and r(t) are the action and reward at the *t* th trial. The parameter α_{Ri} is the learning rate for the rewarded selected action, α_{URi} is the learning rate for the unrewarded selected action and λ_1, λ_2 are the forgetting terms for unselected action rewarded and unrewarded consecutively.

Using the action values, the prediction of the choice at trial t was given by the following:

$$P(a(t) = L) = \frac{1}{1 + e^{-\beta(Q_L(t) - Q_R(t))}}$$

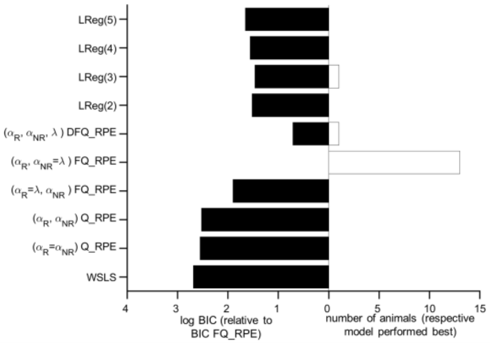

Figure 4: Summary of the different models

Acknowledgments

We thank John-Anthony Fraga and Neil Savalia for their assistance with the collection of data. This research was funded by Kavli Institute for Neuroscience Postdoctoral Fellowship, NIH MH112750, NIH MH118596 and NIH AG047270.

References

Corrado, G., & Doya, K. (2007). Understanding neural coding through the model-based analysis of decision making. *Journal of Neuroscience*, 27(31), 8178-8180.

Ito, M., & Doya, K. (2009). Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *Journal of Neuroscience*, *29*(31), 9861-9874.

Makino H, Ren C, Liu H, Kim AN, Kondapaneni N, Liu X, Kuzum D, Komiyama T. Transformation of Cortex- wide Emergent Properties during Motor Learning. Neuron. 2017;94(4):880-90 e8.

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. Annual Review of Neuroscience, 35, 287-308.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, 22(6), 1027-1038.

Seo H, Lee D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J Neurosci. 2007;27(31):8366-77.

Siniscalchi MJ, Phoumthipphavong V, Ali F, Lozano M, Kwan AC. Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior. Nat Neurosci. 2016;19(9):1234-42.

Sul JH, Jo S, Lee D, Jung MW. Role of rodent secondary motor cortex in value-based action selection. Nat Neurosci. 2011;14(9):1202-8.