Updating And Maintenance Dynamics Of Action Values During Foraging Behavior

C. Barezzi[1], S. Fuentes[1], B. Serrano-Porcar [1], R. Marín-Campos [1], M. Molano-Mazón[1], C. Sindreu[1], J. De La Rocha[1]

[1] Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, 08036, Spain.

During natural foraging behavior subjects use previous choices and their outcomes to estimate the changing value of available options, and use these action values to guide future decisions. Despite recent progress in the study of rodent foraging (Bari *et al*. 2019; Hattori *et al*. 2019), it remains unclear how neural circuits update and maintain these action values and which are the brain areas involved in this process.

Here, we investigated how mice performed a two-armed bandit (2AB) task in which they chose between two options that provided reward with different probabilities. The reward probabilities kept changing across trial blocks of different duration (mean 30 trials) so that mice had to track the value of the options across trials. To investigate the maintenance of these values across time, Inter-trial-intervals (ITIs) were drawn from an exponential distribution (rate parameter of 0.5, maximum of 30 s), during which animals had to wait to initiate the new trial. We used a fully automated system, the Mouse Village (MV), that permitted mice to self-train by freely choosing their working routine, without any human intervention.

Animals showed task adaptive foraging behavior in which their choice probabilities mirrored the time-varying reward probabilities. Logistic regression analysis indicated that mice choices tended to repeat previous rewarded responses, but were surprisingly unaffected by non-rewarded ones. Moreover, choice probabilities were not affected by ITIs revealing the stability of the encoding of action values. We trained Recurrent Neural Networks (RNNs) on the same 2AB task and found that they outperformed mice by using both previous rewarded and unrewarded responses to update action values and guide choice behavior.

Together, our results show that animals developed a suboptimal strategy in which action values were remarkably maintained over long waiting periods, but were updated based on only previous positive outcomes. Future work will employ optogenetic targeting of medial prefrontal cortex and dorsomedial striatum to probe their role in value updating and maintenance during foraging behavior.