I. Motivation

Decision-making is a critical skill for animals and autonomous robots alike. Whether you are a rabbit or a driverless car, you constantly need to make appropriate decisions. This work stresses the importance of taking into account habit formation in decision-making and in goal-directed behaviors such as intrinsic motivation, especially as it pertains to sensorimotor learning.

In computational systems, reinforcement learning (RL) (Sutton, 1998) has been a popular framework to describe and explore the issues of decision-making. Interestingly, although the RL framework was not intended to provide a plausible model of reinforcement learning in animals, RL, and in particular temporal difference (TD), has been a popular choice to model and explain observed experimental data in neuroscience (Pan, 2005); this is in great part due to TD providing a temporal model for the Rescorla and Wagner learning rule (Rescorla, 2005). The reward prediction error of TD has been a popular choice to model the Rescorla and Wagner learning rule, and the one from the cognitive reinforcement learning rule, and the one from the cognitive

II. Computational Model

Our lab previously created a neurocomputational model of the basal ganglia (Topalidou et al., 2016). The model is implemented as a recurrent neural network with rate-coded neurons reproducing the main structures and interactions found in the basal ganglia. The network is organized as three interactive loops: motor, associative and cognitive. The cognitive loop perceives the visual stimuli, and the motor loop generates the action that pushes the chosen button, and the associative loop encodes the mapping between stimuli and buttons.

Out of all synaptic connections in the network, only two are plastic. The connection between the cognitive cortex and the cognitive striatum changes its weights according to a reinforcement learning rule, and the one from the cognitive to the associative cortex implements Hebbian learning. While the former is affected by rewards, the latter is not.

We subjected this model to a two-armed bandit task where two visual stimuli, A and B, are presented. Stimuli are rewarded probabilistically, with probability $r_A$ and $1 - r_A$ respectively: if $r_A = 0.8$, A is rewarded 8 out of 10 times, while B only 2 out 10. For the first 20 trials however, we forced the model to choose a stimulus by presenting only one.

During this period, A and B are presented with a ratio $P_A$ and $1 - P_A$ respectively. For instance, with $P_A = 0.3$, A and B are presented alone, as forced choices, 6 and 14 times respectively in a random order during the first 20 trials. During the rest of the trials, both A and B are presented to the model.
A reward

0

20%

A chosen

B chosen

FORCED FREE CHOICES

Average stimulus choice for 100 repetions of the task.

A and B are rewarded 75% and 25% of the time respectively.

Fig. 2. In the model, choices can go against rewards, if the stimulus with the lower has been sufficiently reinforced enough through Hebbian learning. In this instance, \( r_A = 0.75 \) and \( P_A = 0.2 \). The figure shows averages over 100 runs. Despite A being much more rewarded than B, B is initially preferred when free choices are allowed. The choice then reach a balanced equilibrium: around 50 runs have switched to always choosing A (RL was stronger), the other to always choosing B (Hebbian learning was stronger). The behavior of the model is heavily affected by the relative weights of Hebbian learning and RL, as well as the initial RL value for the two stimuli. The parameters were fitted on experimental data from previous monkey experiments.

which is able to choose freely between them. Our hypothesis is that if we force the choice of the less rewarded stimulus B sufficiently more often than A, the model will choose B more than A when able to choose freely.

As shown in Figure 2, the model is indeed able to display such a behavior. The interpretation is that during the forced choice, the Hebbian connection is strongly reinforced towards B. In other words, the model habituates to choosing B: the more a choice is made, the easier it is to make it in the future. Under some circumstances (see Figure 3), it allows Hebbian learning to prevail over reinforcement learning. For instance, if B is chosen all the time over a time period, A is not reinforced through RL anymore while Hebbian influence favoring B grows, making choosing A even less likely in the future. Interestingly, whereas novelty-based intrinsic motivation models predict that (moderately) novel stimuli are more attractive, our model explores an opposite tendency: the decision process favors familiarity of choice.

III. BEHAVIORAL EXPERIMENTS

To test the predictions of the model, we are applying the two-armed bandit protocol to non-human primates (macaca mulata, see Figure 4). Unlike the model, during the forced choice period, A and B do not appear alone. They appear with a neutral stimulus as an alternative; the neutral stimulus is never rewarded. Additionally, the forced choice period last 50 trials. The experiments are ongoing.

IV. ACKNOWLEDGEMENTS

We wish to thank Hugues Orignac, Tho-Hai Nguyen, Aurélien Nioche and Brice de la Crompe for their support, advice, and insights.

Source code and data for the figures can be found at: http://dx.doi.org/10.6084/m9.figshare.5203993

REFERENCES


Topalidou, Meropi et al. (2016). Dissociation of reinforcement and Hebbian learning induces covert acquisition of value in the basal ganglia. Tech. rep. doi:10.1101/060236


1Experimental procedures were performed in accordance with the Council Directive of 20 October 2010 (2010/63/UE) of the European Community. This project was approved by the French Ethic Comity for Animal Experimentation (50120111-A).