Journal Club

Editor's Note: These short reviews of a recent paper in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to mimic the journal clubs that exist in your own departments or institutions. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Simple Reinforcement Learning Models Are Not Always Appropriate

Hansem Sohn and Seungyeon Kim

Brain Dynamics Laboratory, Department of BioSystems, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea

Review of Hampton et al. (http://www.jneurosci.org/cgi/content/full/26/32/8360)

Approximately a decade has passed since midbrain dopamine neurons emerged as a bridge between research in traditional machine learning, particularly reinforcement learning (RL), and behavioral neuroscience (Schultz et al., 1997). In RL, the difference between the actual reward and the expected reward is called the "prediction error," which is used in learning models to optimize the future reward. Dopamine neurons in the ventral tegmental area and substantia nigra are involved in the error prediction. Multiple reward signals are sent to brain regions involved in sophisticated decision-making processes, such as basal ganglia, the anterior cingulate cortex (ACC), and the prefrontal cortex (PFC). However, simple RL models do not encompass all aspects of elaborate human decision-making processes. For example, the interdependencies in states, actions, and ensued rewards are not characterized in simple RL models because they only update the selected option. This disadvantage can be overcome by model-based RL in which the dynamics of the given conditions are learned indirectly by the construction of an abstract model that includes the structure of particular tasks.

In their recent *Journal of Neuroscience* article, Hampton et al. (2006) aimed to determine whether human subjects engage in a simple decision-making problem by using state-based inferential knowledge of the task structure or by a simple RL model based on the individual reward history. Sixteen subjects were scanned with functional magnetic resonance imaging while performing a probabilistic reversal learning task that incorporated anticorrelation between the reward distributions associated with two choices and the knowledge that the contingencies will reverse [Hampton et al. (2006), their Fig. 1*C* (http://www.jneurosci.org/cgi/content/full/26/32/8360/F1)]. Specifically, on each trial, subjects were presented with two stimuli designated as the correct and incorrect choice. After subjects selected a stimulus, a monetary gain or loss indicative of a reward or punishment, respectively, appeared on the screen in a stochastic manner. The contingencies also reversed randomly after four consecutive correct choices. When the reversal occurred, subjects needed to select a new correct stimulus before another reversal took place.

Hampton et al. (2006) implemented a Bayesian hidden Markov model for their task. The model statistically inferred the probability of being in the correct-choice state from the choice action (stay or switch), the observed reward, and the probability of previous choice states. Their model was compared with simple RL models, such as Q-learning, actor-critic methods, and advantage-learning models. At the behavioral level, the state-based model produced more accurate fitting with choices of subjects compared with all simple RL models. Thus, subjects managed the probabilistic reversal learning with implicit integration of task structure. To identify the neural substrates underlying the task, the random-effects regression analysis of fMRI data were performed with the estimated state-based model and best-fitting RL model. The prior–correct probability of the state-based model, which is the expected value in the simple RL models, was represented in the medial PFC, orbitofrontal cortex, and amygdala. In addition, the ventral striatum, dorsomedial PFC, and ventromedial PFC (vmPFC) encoded a posterior–prior correct update signal analogous to the prediction error in the simple RL models [Hampton et al. (2006), their Fig. 2*D* (http://www.jneurosci.org/cgi/content/full/26/32/8360/F2)]. The ingenuity of the experimental design revealed both the similarity and the difference in the models. Surprisingly, the neural activity patterns of the vmPFC followed the prediction of the state-based model that is specifically distinct from that of the simple RL after subjects switched their choice in the punished trial [Hampton et al. (2006), their Fig. 3*A* (http://www.jneurosci.org/cgi/content/full/26/32/8360/F3)]. In simple RL, the value of the chosen option is updated, but the other option will not be updated until it is chosen. Thus, the decision to switch to the new option will result in a less expected value. However, the decision

Received Sept. 12, 2006; revised Sept. 19, 2006; accepted Sept. 19, 2006.

Correspondence should be addressed to Seungyeon Kim, Brain Dynamics Laboratory, Department of BioSystems, Korea Advanced Institute of Science and Technology, Daejeon 305-701 Republic of Korea. E-mail: seungyeonkim@kaist.ac.kr.

DOI:10.1523/JNEUROSCI.3973-06.2006

 0270-6474/06/2611511-02$15.00/0

to switch in the state-based model has shown that the expected value is high, as reflected in neural activities in the vmPFC.

In our everyday lives, the very same notion prevails. After "plan A" fails, we switch to "plan B," hoping for a more positive outcome with the renewal of prior probability for plan B. The greater probability of success resides even before we execute plan B and validate the successfulness of it. Hampton et al. (2006) designed a model-based RL model that can effectively depict the human goal-directed behavior in such situations. Furthermore, they identified the vmPFC, ACC, and anterior insular as the neural underpinnings of stated-based decision making. Combined with a recent study that the vmPFC is involved in avoidance learning (Kim et al., 2006), these results indicate that the vmPFC has a variety of roles in decision-making situations, along with other brain regions.

This study provides a step in the right direction, but it reflects only the tip of the iceberg. First, the number of states and corresponding actions is likely to be more than two, and the degree of interdependency might be varied. Even for infinite possible actions (e.g. continuous-valued actions), simple RL models, such as actor-critic methods, might be superior to the state-based model because of computational load (Sutton and Barto, 1998). Second, heterogeneous stimuli that are in appetitive, monetary, and social rewards are thought to be differentially processed in our brain. Third, our decision-making processes have dynamic flexibilities that are neglected in the state-based model. The stationary model possesses constant parameters and the Markov property. In reality, decision models are dynamic in a way that the sequence of the previous outcomes influences the calculation of the probability of being correct or incorrect.

There are multiple neural systems that govern human decision making (Daw et al., 2005). Hampton et al. (2006) show that simple RL and abstract state-based models make qualitatively different predictions in different neural systems and that simple RL models are not always appropriate for human decision making. Without a doubt, the RL with midbrain dopamine neurons is at the core of multiple learning models and goal-directed behavior. The convergent and divergent interactions between neural systems seem to result in complex decision making. Future longitudinal investigations should concentrate on experimental and theoretical work to unravel the intricate human decision-making processes that may involve different learning models in various neural systems concurrently and sequentially.

References

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711.

Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci 26:8360–8367.

Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol 4:e233.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.

Sutton RS, Barto AG (1998) Reinforcement learning. Cambridge, MA: MIT.