Decisions about Decisions

Wako Yoshida[1,*] **and Ben Seymour**[2,3,*]
[1]Advanced Telecommunications Research Institute International, Kyoto 619-0288, Japan
[2]Center for Information and Neural Networks, National Institute of Information and Communications Technology, Osaka 565-0871, Japan
[3]Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
*Correspondence: yoshida@atr.jp (W.Y.), bjs49@cam.ac.uk (B.S.)
http://dx.doi.org/10.1016/j.neuron.2014.01.030

A major puzzle of decision making is how the brain decides which decision system to use at any one time. In this issue of *Neuron*, Lee et al. (2014) provide a theoretical, behavioral, and neurobiological account of a prefrontal reliability-based arbitration system.

As sure as there are many ways to skin a proverbial cat, there are many ways to solve most real-life decision-making problems. That the brain has several different learning and decision systems at its disposal is no longer disputed but has given way to a much trickier question: how do you decide which decision system to use at any one time?

In its simplest instantiation, this "meta-decision" problem can be thought of as a choice between a computationally extravagant "model-based" system that tries to build a full internal model of the external world, and a frugal "model-free" system that adopts a "what usually works well" approach. Given that much in the world is actually rather mundane and predictable, this latter system of habits will easily suffice for the majority of the time (Dolan and Dayan, 2013). The problem is knowing when it is sufficient and safe to rely on it.

In 2005, Daw and colleagues presented the first specific computational account of how such arbitration might be controlled: they suggested that as well as outputting their preference (i.e., values) of possible actions, each system might also accompany this with an estimate of the uncertainty in these values (Daw et al., 2005). This uncertainty signal could provide a normative basis for arbitration, allowing optimal weighting of the values outputted by each system to allow an integrated decision. The Daw model provides a good account of the existing data in animals and humans, but it is not particularly easy to test rigorously. Indeed, only recently have paradigms been developed that reliably disambiguate different values produced by each system, and it is not trivial to refine these paradigms so that they also independently vary the uncertainty in these outputs.

In this issue of *Neuron*, Lee, Shimojo, and O'Doherty present a comprehensive theoretical, behavioral, and neurobiological analysis of the arbitration problem (Lee et al., 2014). Their investigation centers around three central questions. First, is there a dynamic, flexible process that arbitrates the respective contributions? Second, if so, what is the key signal or signals that each system outputs to an arbitration module to allow arbitration decisions to be computed? And third, how is such a system implemented in the brain?

Lee et al. (2014) use a combination of instruction and task complexity to independently manipulate values and uncertainties in model-free and model-based systems. Specifically, subjects engage in a two-step decision task in which they make right/left choices to move first to an intermediate state and second to an outcome state that yields some amount of monetary reward signaled by colored coins. The task is performed under two conditions: in the first ("flexible") condition, subjects can cash in coins of any color: this makes the task relatively easy, indeed easy enough for a model-free system, as simply reinforcing action values works well. In the second ("specific") condition, only coins of a given color can be cashed, with the others being worthless. This favors a model-based system, which can plan the best action by memory of the particular color of the coins in each of the outcome states. To further flexibly manipulate the uncertainty (noting also that it is still possible that a model-free learning system could learn the task with an expanded representation of the initial state), Lee et al. (2014) also manipulate the probability that a given action yields each subsequent state, to be either high (0.9) or low (0.5). Importantly, these different state-transition probabilities differentially affect the specific and flexible conditions and induce a dynamic variability in uncertainty within a temporal range detectable with fMRI.

In considering how the balance of control might shift between systems, Lee et al. (2014) propose a new model of arbitration (Figure 1), which although similar to Daw's uncertainty-based model, has two important differences. First, to evaluate the system's prediction, they used the reliability, the variance-to-mean ratio of the probability that the prediction error is zero at a moment, instead of the variance per se (uncertainty in Daw et al.) or the mean of prediction error. Second, rather than using trial-by-trial reliabilities to instantaneously determine the relative contribution of each system, they propose a dynamical two-state transition model, in which the reliabilities modulate the transition rate between choice probabilities of the two systems. This yields a gradual shift in the reliance on either system, as opposed to a knee-jerk dependency. The fact that the habits tend to emerge with increased training, previously represented as an exponential decay (Gläscher et al., 2010), was accommodated by a bias on the transition rates so that model-free control is favored if the reliabilities are equal. In the behavioral data, model fitting suggests that subjects' choices seem to better reflect incorporation of both these differences.

Neurobiologically, they find that reliability signals relating to both systems have an overlapping representation in an

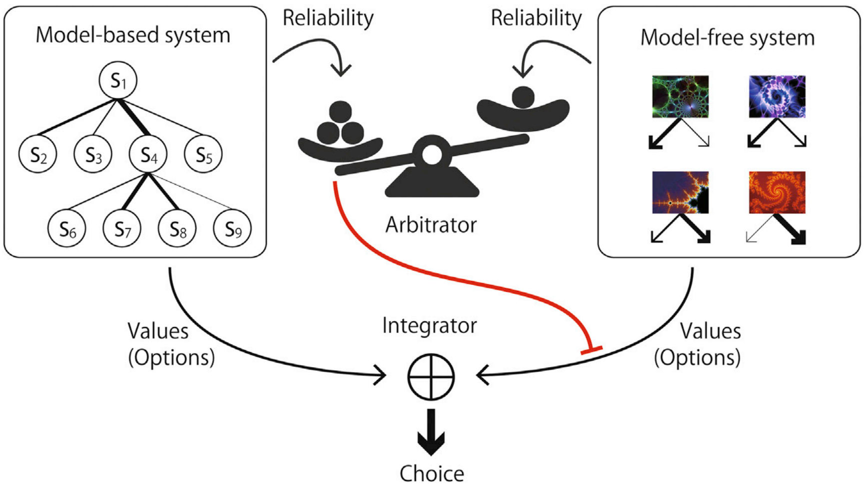

Figure 1. Schematic Diagram of Arbitration
The model-based system learns the structure of state space, including transition probabilities and reward, whereas the model-free system simply learns action values. Each system outputs a set of action values to an integrator, and a reliability signal relating to these action values to the arbitrator. Stronger reliabilities shift the arbitration balance (which has a slight bias toward the model-free system) in favor of that system, which yields control by flexible inhibition of the model-free system.

anterior region of lateral inferior prefrontal cortex, with this region and a region of frontopolar cortex correlating best with the reliability of the more reliable system at any one time. This suggests that this region could be involved in subsequent arbitration computations. Interestingly, effective connectivity was modulated by the arbitrator's output: if the arbitrator favored the model-based system, there was increased negative coupling between the lateral inferior prefrontal cortex and the putamen (implicated in the model-free system). This suggests that the reliability-based arbitrator might inhibit "default" model-free system to favor model-based control rather than controlling the two systems symmetrically.

What is appealing about these results is that they provide a seemingly direct demonstration of a natural hierarchy in the control of decision making, with a decision "metacontroller" exerting influence over "lower" individual decision systems. This is a self-organizing hierarchy, with the systems themselves providing the information that is used by the arbitrator, so in principle the metacontroller, as the apical node, does not necessarily require any additional information—it just dishes out control in a principled manner. This makes it the first neurobiological observation of a complete multisystem decision process.

In addition to suggesting a new role for the anterior region of the inferior lateral prefrontal cortex, the results shed light on the function of the frontal pole—an area that seems to subserve some of the most intriguing but obscure functions in the human brain. Broadly speaking, the frontopolar cortex appears to compute relations among internally maintained contextual representations (of which reliability may be just one type), to contribute to the flexible updating of behavior in dynamic environments. Neuroimaging studies of decision making have shown responses associated with nonpreferred options (Koechlin et al., 1999), lesser-valued exploratory options (Daw et al., 2006), and "next best" alternatives (Boorman et al., 2009). It has also been shown that activity seems to track the relative advantages of options within one control system, i.e., uncertainties of predictions from model-free (Badre et al., 2012) and model-based (Yoshida and Ishii, 2006) systems. However, the current study is notable as it is the first to demonstrate reliability-like signals in multiple learning systems.

As a result, as Lee et al. (2014) speculate, this may allow the frontopolar cortex to supervise the inhibitory control in the lateral inferior prefrontal cortex as a "controller of controllers." However, this still leaves open the question of adaptive control, i.e., whether this region learns to control (for example, by building more abstract representations) or merely implements control. Both inherently noisy systems and imperfect or partial observability make the information from environments uncertain, but the latter type of uncertainty can be reduced by belief inference—a posterior probability over possible options. In Lee et al. (2014)'s task, for example, the probabilistic state transition is systems uncertainty, but there is also a higher state uncertainty regarding the existence of the distinct uncertainty conditions (0.5 or 0.9, which is not known to the subjects). Thus, there is the potential for more complex, hierarchical inference-based model representation and that may have its own distinct uncertainty signal and influence on the control policy.

The complexity of the decision making means that the space of subtle differences in the structure of the model-based system, input functions, arbitration mechanism, and output influence is large. This creates the opportunity for extensive debate on the precise details of what is being computed in similar tasks, which can easily become complicated by the methodological challenges of fitting multiple similar models and potentially codependent parameters. Of course, it is always easy to make the model more complicated, but this should detract from the limpidity and fecundity of the current exposition.

However, there are some extensions that are likely to be especially interesting. For example, how do Pavlovian values exert competitive control (i.e., during Pavlovian-instrumental interactions)? Is there a flexible, parametric influence of the computational "effort cost" associated model-based processing (which is fixed in the current model)? It is also worth noting that there are other potentially interesting additional ways in which model-based and model-free systems can interact. First, it is possible that each system could take advantage of the prediction errors generated by the other. Second, when it comes to choice, the model-based system might have access to model-free values when planning (the model-based systems' internal representation might include the model-free system). Third, accumulated control by the model-free system might ultimately inhibit not just planning but also the learning of the model-based system.

Notwithstanding this, it will be fascinating and illuminating to establish to what extent these findings will generalize to other types of task, especially those that place different demands on internal representations inherent in the model-based system. This is important because almost certainly there are different types of model-based learning and planning system: for example, rule- or instruction-based models (as explored here), model-based avoidance, partially observable Markov decision problems (Yoshida and Ishii, 2006), hierarchical decision-making problems (Ribas-Fernandes et al., 2011), and navigation (Simon and Daw, 2011). This puts the onus on other groups to emulate the sophisticated modeling of behavior and brain illustrated here.

REFERENCES

Badre, D., Doll, B.B., Long, N.M., and Frank, M.J. (2012). Neuron *73*, 595–607.

Boorman, E.D., Behrens, T.E.J., Woolrich, M.W., and Rushworth, M.F.S. (2009). Neuron *62*, 733–743.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Nat. Neurosci. *8*, 1704–1711.

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Nature *441*, 876–879.

Dolan, R.J., and Dayan, P. (2013). Neuron *80*, 312–325.

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). Neuron *66*, 585–595.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., and Grafman, J. (1999). Nature *399*, 148–151.

Lee, S.W., Shimojo, S., and O'Doherty, J.P. (2014). Neuron *81*, this issue, 687–699.

Ribas-Fernandes, J.J.F., Solway, A., Diuk, C., McGuire, J.T., Barto, A.G., Niv, Y., and Botvinick, M.M. (2011). Neuron *71*, 370–379.

Simon, D.A., and Daw, N.D. (2011). J. Neurosci. *31*, 5526–5539.

Yoshida, W., and Ishii, S. (2006). Neuron *50*, 781–789.