

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Neural Mechanisms for Undoing the “Curse of Dimensionality”

 Avinash R. Vaidya

Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, H3A 2B4, Canada
Review of Niv et al.

Human behavior is marked by a sophisticated ability to attribute outcomes and events to choices and experiences with surprising nuance. Understanding the mechanisms that govern this ability is a major focus for cognitive neuroscience. Reinforcement learning (RL) theory has provided a tractable framework for researching this process and interpreting putative neurophysiological signals underlying learning. The Rescorla–Wagner model in particular (Rescorla and Wagner, 1972), extended into temporal difference learning by Sutton and Barto (1998), has provided an especially rich set of predictions that align nicely with many behavioral and physiological results (Schultz et al., 1997). At the core of these models is the idea that the value of available options is continuously updated by comparing their expected value with feedback after each decision. This comparison yields a prediction error that is used to update expectations and guide future choices. While this model provides an elegant explanation for learning in many simple experimental conditions, it cannot be easily applied to more complex tasks, particularly when options have multiple features,

or dimensions that may each have some value. This problem is even more obvious in real-life choices, where options are so multifaceted and multidimensional that a RL process would seem implausible. This problem has been described as the “curse of dimensionality” (Sutton and Barto, 1998).

The problem of learning in multidimensional environments has long vexed animal learning theorists. Early experiments demonstrated that associative learning did not operate equally for all stimuli: Pavlov (1927) showed that conditioning by a salient stimulus could overshadow learning about a concurrent, less-salient stimulus. Related behavioral phenomena prompted learning theorists to develop more sophisticated models that included attentional modules that would adaptively select features for further learning (Mackintosh, 1975; Pearce and Hall, 1980). These models are linked by a common focus on the importance of learning about the predictive or information value of stimulus dimensions. While the predictions of these models have been investigated thoroughly through years of animal experiments (for review, see Le Pelley, 2010), the underlying neurobiological mechanisms of these processes are not well understood, particularly in the human brain.

In a recent article in *The Journal of Neuroscience*, Niv et al. (2015) investigated the neural correlates of learning in a multidimensional environment, using a combination of computational modeling

and functional magnetic resonance imaging (fMRI). They describe a putative computational mechanism for adaptively selecting relevant features for learning and a network of brain regions that may be involved in this process. In their experiment, subjects chose between three compound stimuli that were each defined by three different dimensions (shape, texture, and color). Subjects completed a series of short blocks, in which only one dimension of the stimuli (e.g., shape) was predictive of rewarding feedback, with rewards being probabilistically more likely for one feature (e.g., 75% chance of reward for the triangle), while the other two features in that dimension were associated with a lower likelihood of reward (e.g., 25% chance of reward for the circle or square). Niv et al. (2015) argue that this task can be solved through representation learning, i.e., electing the current state representation (or relevant stimulus dimension) in the task to guide learning.

The authors compared six computational models that fell broadly into three categories: RL models based on the Rescorla–Wagner equation, a statistically optimal Bayesian model, and a serial hypothesis model that selected between candidate stimulus dimensions based on available evidence. The data were best fit by a reinforcement-learning model operating on the level of stimulus features and incorporating a decay parameter that weakened features not chosen in each trial (fRL+decay model). The authors then asked where the hemodynamic response

Received June 25, 2015; revised July 30, 2015; accepted July 31, 2015.

This work is supported by a CIHR operating grant (MOP 97821), and a Desjardins Outstanding Student Award to A.R.V.

The author declares no competing financial interests.

Correspondence should be addressed to Avinash R. Vaidya, Montreal Neurological Institute, McGill University, 3801 University Street, Room 276, Montreal, QC, H3A 2B4, Canada. E-mail: avinash.vaidya@mail.mcgill.ca.

DOI:10.1523/JNEUROSCI.2428-15.2015

Copyright © 2015 the authors 0270-6474/15/3512083-02\$15.00/0

correlated with representation learning, as measured by the standard deviation of feature weights predicted by the fRL+decay model. This analysis identified a bilateral network of regions corresponding to the frontoparietal attention network described by Corbetta and Shulman (2002), including the intraparietal sulcus (IPS) and dorsolateral prefrontal cortex, as well as the right lateral orbitofrontal cortex (OFC).

At first glance, it may seem surprising that the fRL+decay model described in this work provided the best fit to participants' choices. As mentioned earlier, it seems implausible that RL would operate on all available stimulus dimensions in such a task. Indeed, prior work with this task identified serial hypothesis testing as the best explanation of participants' behavior (Wilson and Niv, 2012), yet this model was outdone by fRL+decay here. The decay parameter of the fRL+decay model was critical to its performance. By decaying unchosen feature weights in every trial, the model effectively puts an attentional filter on features not chosen while still giving them access to the decision process. This relatively simple modification of the RL model made a substantial difference: without the decay parameter, a simple feature-based RL model was marginally outperformed by serial hypothesis testing. These results suggest that simple attentional mechanisms operating in the feature space of such a task provide a better explanation of associative learning than serial hypothesis testing or feature-based RL without an attention mechanism.

In all of the models tested by Niv et al. (2015), feature selection occurs during the decision-making stage. However, earlier attentional RL models suggested that attentional weights are applied to the prediction error signal itself. These earlier models also suggested that teaching signals are weighted based on the learned predictive value of a stimulus feature, rather than through decaying the weights of unchosen features (Mackintosh, 1975; Pearce and Hall, 1980). Niv et al. (2015) do not test this possibility, though they do allude to it in the discussion. Formally comparing these models with the fRL+

decay model would give some insight into the stage at which attentional processes operate during multidimensional learning.

Niv et al. (2015) do not attempt to distinguish the different contributions of regions identified by their analysis to representation learning. There remain many questions about how RL mechanisms interface with attentional selection to guide behavior. Recent work suggests that the IPS signals the behavioral relevance of options during decision-making (Peck et al., 2009). Hunt et al. (2014) have suggested that this region might select attributes based on behavioral relevance, showing that communication between IPS, lateral OFC, and putamen depended on the respective relevance of stimuli and actions to a current choice. Like Niv et al. (2015), these findings point to a role for IPS and lateral OFC in representation learning and selection among stimulus features.

There has recently been a great deal of interest in the role of OFC in model-based RL. It has been suggested that this region provides a cognitive map that allows representation of the underlying reinforcement contingencies of a task (Wilson et al., 2014; Stalnaker et al., 2015). The experiment in Niv et al. (2015) similarly requires adaptively learning which stimulus features are currently relevant, demanding that subjects quickly learn the hidden state of reinforcement in each game. OFC might operate together with lateral prefrontal cortex and IPS to adaptively attend to relevant stimulus features based on a model of the reinforcement contingencies in the task. Future work, possibly using techniques with a faster time resolution and better signal quality in OFC will be required to understand how this network of regions operates together to guide representation learning.

The results of Niv et al. (2015) provide an important step forward in understanding the neurobiological and computational mechanisms underlying representation learning. The authors demonstrate that some straightforward modifications of the basic RL model can greatly improve its predictive power in a fairly complex task. More broadly, these results point a way forward for using the relatively simple

mechanics of RL for generating tractable computational hypotheses for learning in complex environments. This line of work might prove useful as researchers move toward grappling with understanding the neurobiology of behavior in more ecologically valid settings.

References

- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215. [CrossRef Medline](#)
- Hunt LT, Dolan RJ, Behrens TE (2014) Hierarchical competitions subserving multi-attribute choice. *Nat Neurosci* 17:1613–1622. [CrossRef Medline](#)
- Le Pelley ME (2010) Attention and human associative learning. In: *Attention and associative learning: from brain to behaviour* (Mitchell CJ, Le Pelley ME, eds.). Oxford: Oxford UP.
- Mackintosh NJ (1975) Theory of attention-variations in associability of stimuli with reinforcement. *Psychol Rev* 82:276–298. [CrossRef](#)
- Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, Radulescu A, Wilson RC (2015) Reinforcement learning in multidimensional environments relies on attention mechanisms. *J Neurosci* 35:8145–8157. [CrossRef Medline](#)
- Pavlov IP (1927) *Conditioned reflexes*. London: Oxford UP.
- Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87:532–552. [CrossRef Medline](#)
- Peck CJ, Jangraw DC, Suzuki M, Efem R, Gottlieb J (2009) Reward modulates attention independently of action value in posterior parietal cortex. *J Neurosci* 29:11182–11191. [CrossRef Medline](#)
- Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: *Classical conditioning. II. Current research and theory* (Black AH, Prokasy WF, eds.), pp 64–99. New York: Appleton-Century-Crofts.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599. [CrossRef Medline](#)
- Stalnaker TA, Cooch NK, Schoenbaum G (2015) What the orbitofrontal cortex does not do. *Nat Neurosci* 18:620–627. [CrossRef Medline](#)
- Sutton RS, Barto AG (1998) *Reinforcement learning: an introduction*. Cambridge, MA: MIT.
- Wilson RC, Niv Y (2012) Inferring relevance in a changing world. *Front Hum Neurosci* 5:189. [CrossRef Medline](#)
- Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81:267–279. [CrossRef Medline](#)