

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.

A Role for the Human Substantia Nigra in Reinforcement Learning

Archy O. de Berker^{1,2} and Robb B. Rutledge^{2,3}

¹Sobell Department of Motor Neuroscience and Movement Disorders, ²Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom, and ³Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, United Kingdom

Review of Ramayya et al.

In a world rich with stimuli and potential actions, organisms must learn which objects are rewarding and which actions produce rewards. Dopamine neurons may play a key role in learning the values of stimuli and actions by representing reward prediction errors (RPEs), the difference between experienced and predicted reward. Although many studies report that phasic activity of dopamine neurons in the ventral tegmental area (VTA) and substantia nigra (SN) represents RPEs in humans (Zaghloul et al., 2009) and other animals (Schultz et al., 1997; Cohen et al., 2012), only recently has evidence emerged to support an instrumental role for phasic dopamine in reinforcement learning (Steinberg et al., 2013). However, it remains unclear whether the phasic activity of dopamine neurons in the VTA and SN play different roles in reinforcement learning. This issue is of particular importance given the selective deterioration of

SN dopamine neurons in Parkinson's disease (Kish et al., 1988).

Ramayya et al. (2014) recently reported the results of a study addressing this important question. During deep brain stimulation (DBS) surgery to treat Parkinson's disease, patients routinely undergo recording and microstimulation of SN neurons to aid surgeons with DBS electrode placement in the nearby subthalamic nucleus. These operations provided the opportunity to record SN neurons during a probability-learning task and to use electrical microstimulation of those neurons to manipulate behavior.

In three blocks of 50 trials each, subjects made 25 choices with one stimulus pair and 25 choices with another stimulus pair, with the two pairs presented in alternating trains of three to six trials. One stimulus in each pair was associated with a high probability of reward (the "high-probability stimulus") and the other stimulus was associated with a low probability of reward (the "low-probability stimulus"). Importantly, the left–right configuration of the stimuli in each pair was random, and therefore successful task performance required learning the values of stimuli and not the values of actions (i.e., left or right button presses).

In the first block, subjects on average made correct responses in 63% of trials. At the same time, the authors recorded the waveform and phasic spike response to positive feedback for a single SN neuron

in each subject. In the second block, phasic microstimulation was applied coincident with all positive feedback resulting from the high-probability stimulus for one of the two stimulus pairs (the STIM^{POS} pair); this outcome, as a positive RPE, should be associated with an increase in phasic dopamine (Hart et al., 2014). Thus, microstimulation is expected to further increase phasic dopamine release associated with the positive RPE. In the third and final block, phasic microstimulation was applied coincident with all negative feedback from the low-probability stimulus for one of the two stimulus pairs (the STIM^{NEG} pair); this outcome, as a negative RPE, should be associated with a decrease in phasic dopamine (Hart et al., 2014). Thus, microstimulation might counteract the phasic dopamine decrease normally associated with a negative RPE.

The authors reported that correct task performance decreased for the STIM^{POS} stimulus pair, suggesting that SN microstimulation does not enhance the learning of stimulus values needed to perform the task. This immediately suggests that the SN plays a different role in reinforcement learning to the VTA, where microstimulation potentiates stimulus–reward learning (Steinberg et al., 2013; Arsenault et al., 2014). To test whether this decline in performance related to an increased emphasis on learning of action–reward rather than stimulus–reward associations, the authors developed a hybrid action–stimu-

Received July 14, 2014; revised Aug. 11, 2014; accepted Aug. 13, 2014.

This work was supported by the Medical Research Council (A.O.d.B.) and the Max Planck Society (R.B.R.). The Wellcome Trust Centre for Neuroimaging is supported by core funding from Wellcome Trust Grant 091593/Z/10/Z. We thank Peter Dayan for helpful comments.

Correspondence should be addressed to either of the following: Archy O. de Berker, Sobell Department of Motor Neuroscience and Movement Disorders, 33 Queen Square, London WC1N 3BG, United Kingdom, E-mail: archy.berker.12@ucl.ac.uk; or Robb B. Rutledge, Wellcome Trust Centre for Neuroimaging, 12 Queen Square, London WC1N 3BG, United Kingdom, E-mail: robb.rutledge@ucl.ac.uk.

DOI:10.1523/JNEUROSCI.2854-14.2014

Copyright © 2014 the authors 0270-6474/14/3412947-03\$15.00/0

lus learning model (Fig. 1A). In this model, stimulus and action values are learned separately and then combined, using a weighting parameter, into an aggregate value for each action–stimulus combination. In model simulations, increasing the weighting parameter to give action values a greater influence on combined values resulted in a decline in STIM^{POS} performance. Model simulations were also consistent with other features of the data, such as a significant correlation across subjects between the decrease in STIM^{POS} accuracy and the probability of repeating the same action (i.e., left or right button press) after positive feedback, a probability that depends on the extent to which action values influence decision making.

The authors also found that task performance did not decrease for the STIM^{NEG} pair, but their model simulations did not offer an explanation for this surprising result. We therefore explored an alternative interpretation of the results using the same hybrid model but with microstimulation increasing the action-value RPE term that SN dopamine neurons might represent, instead of the weighting parameter (Fig. 1A). We generated simulations using the hybrid model with the same parameters ($\alpha = 0.2$, $\beta = 0.2$) and equal weighting for stimulus and action values ($W_A = 0.5$).

Using this approach, we reproduced the main features of the data, including a greater decrease in accuracy for the STIM^{POS} pair (Fig. 1B) than for the STIM^{NEG} pair (Fig. 1C). For example, a stimulation magnitude sufficient to elicit a 12% accuracy decrease for the STIM^{POS} pair led to only a 4% accuracy decrease for the STIM^{NEG} pair. This asymmetry has a simple explanation: negative feedback should normally weaken action–reward associations but stimulation during that feedback might counteract negative action-value RPEs. Paradoxically, small STIM^{NEG} microstimulation magnitudes could actually improve performance in a task where performance depends only on stimulus values because they might reduce the average difference in values between the actions, and thus the influence of action values on choice (Fig. 1C). Our alternative approach also has the benefit of obviating the need for the brain to store combined action–stimulus values. Instead, action values and stimulus values can be combined during the decision process, which may be more parsimonious than updating action–stimulus values using a weighting parameter that is only affected by stimulation at the

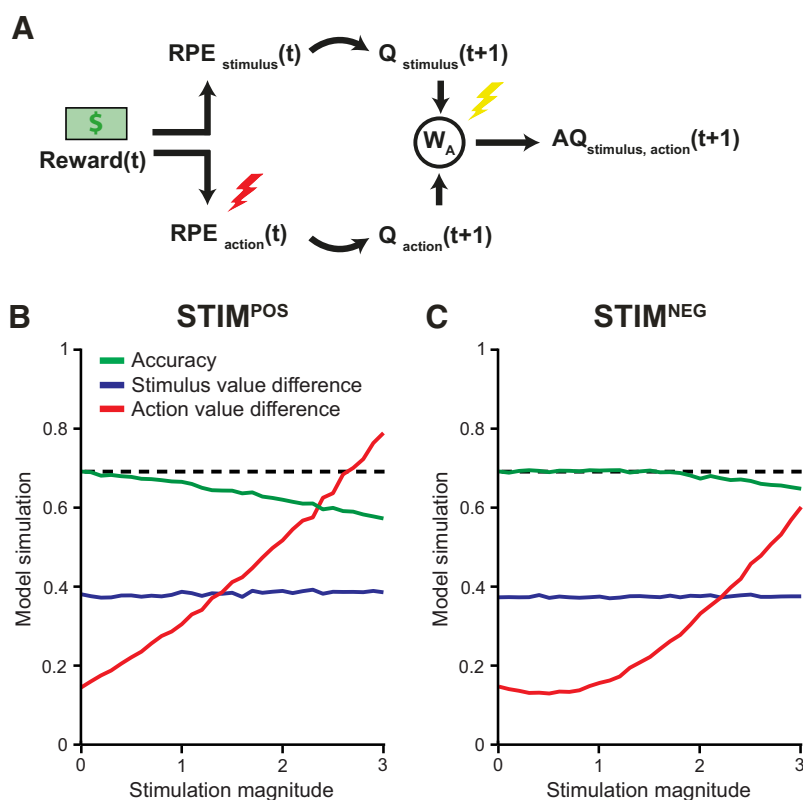


Figure 1. Model simulations for microstimulation effects on reinforcement learning. **A**, The hybrid action–stimulus learning model of Ramayya et al. (2014) features separate stimulus values (Q_{stimulus}) and action values (Q_{action}) that are aggregated with a weighting parameter W_A into combined values ($AQ_{\text{stimulus, action}}$) used to make decisions. Stimulation is suggested to increase W_A (yellow lightning bolt), thereby increasing the influence of action values on decisions. An alternative possibility (red lightning bolt) is that stimulation increases action-value RPEs used to update action values. **B**, In the STIM^{POS} condition, stimulation coincides with positive feedback from the high-value stimulus for the STIM^{POS} pair. If phasic dopamine activity represents action-value RPEs, then as the magnitude of SN dopamine neuron stimulation increases, the average difference in the action values for the two actions increases (red) and action values exert a greater influence on decision making relative to stimulus values (blue). As the influence of action values increases, performance in a stimulus-value learning task decreases (green). **C**, In the STIM^{NEG} condition, stimulation coincides with negative feedback from the low-value stimulus for the STIM^{NEG} pair. For low magnitudes, stimulation does not increase the influence of action values on decision making. Higher stimulation magnitudes are necessary to obtain deficits comparable to the STIM^{POS} condition.

time of outcome delivery. Further research will be needed to determine exactly how the weighting parameter exerts its effect.

The results of Ramayya et al. (2014) suggest that SN dopamine neurons represent action-value RPEs, complementing the possibility that VTA dopamine neurons represent stimulus-value RPEs. This division of labor between the SN and VTA is consistent with distinct patterns of inputs to VTA and SN dopamine neurons (Watabe-Uchida et al., 2012), recent findings that stimulating VTA dopamine neurons increases stimulus–reward associations (Steinberg et al., 2013; Arsenault et al., 2014), and the finding that optogenetic stimulation of VTA and SN dopamine neurons have similar effects on operant place preference in mice when, as in many tasks, stimulus-value or action-value learning make similar behavioral predictions (Ilango et al., 2014).

How might the role of the SN in reinforcement learning be further tested? If the SN plays a role in action-value but not stimulus-value learning, microstimulation of SN dopamine neurons in Parkinson’s patients should improve STIM^{POS} performance in tasks that could be learned by either mechanism, and in tasks in which feedback depends only on actions. Furthermore, electrophysiological recordings during these tasks could reveal whether VTA and SN dopamine activity represent different types of RPEs or reflect the weighting parameter used in different tasks.

A prominent proposal (the actor–critic model) suggests that stimulus-value RPEs represented by dopaminergic projections from the VTA to the ventral striatum (the “critic”) underlie learning of stimuli or states necessary for Pavlovian conditioning. The same stimulus-value RPEs are then used to inform instrumen-

tal choice, training an “actor” putatively located in the dorsal striatum (Balleine et al., 2008). Neuroimaging (O’Doherty et al., 2004) and pharmacological (Piray et al., 2014) studies support this distinction, but it is unclear what the role of SN dopamine neurons might be in this scheme, since electrophysiological results suggest that SN dopamine neurons do not represent stimulus–value RPEs (Morris et al., 2006). The hybrid action–stimulus learning model introduced by Ramayya et al. (2014) differs from the actor–critic model in that values associated with actions can be learned in isolation from stimulus values (although it remains unclear whether this learning is entirely stimulus-independent or whether action values are separately linked to each stimulus pair). If SN dopamine neurons are not used to learn stimulus values, as these findings suggest, SN microstimulation might have no effect on Pavlovian phenomena such as vigor, which might be reflected in reaction times. Further studies might use microstimulation of the VTA and SN to explicitly compare the predictions of these differing models of basal ganglia function.

The finding that changes in accuracy can be understood as changes in the weighting of action and stimulus values raises some interesting possibilities. Early-stage nonmedicated Parkinson’s patients might show deficits in an action–value learning task due to loss of SN dopamine neurons but might actually outperform control subjects in a stimulus–value learning task due to a reduced influence of

action values on their choices. The weighting parameter introduced in this study may also provide an informative way of characterizing behavior, providing a behavioral assay that may relate to patterns of dopamine depletion in Parkinson’s disease.

The demonstration by Ramayya et al. (2014) of a role for the human SN in action–value learning is an important advance in our understanding of reinforcement learning in humans. These results suggest several interesting directions for future research that might clarify the precise roles of VTA and SN dopamine neurons and at the same time advance our understanding of Parkinson’s disease and the effects of dopaminergic drugs on behavior.

References

- Arsenault JT, Rima S, Stemmann H, Vanduffel W (2014) Role of the primate ventral tegmental area in reinforcement and motivation. *Curr Biol* 24:1347–1353. [CrossRef Medline](#)
- Balleine BW, Daw ND, O’Doherty JP (2008) Multiple forms of value learning and the function of dopamine. In: *Neuroeconomics: decision-making and the brain* (Glimcher PW, Camerer C, Fehr E, Poldrack RA, eds.), pp 367–385. New York: Academic.
- Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482:85–88. [CrossRef Medline](#)
- Hart AS, Rutledge RB, Glimcher PW, Phillips PE (2014) Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J Neurosci* 34:698–704. [CrossRef Medline](#)
- Ilango A, Kesner AJ, Keller KL, Stuber GD, Bonci A, Ikemoto S (2014) Similar roles of substantia nigra and ventral tegmental area dopamine neurons in reward and aversion. *J Neurosci* 34:817–822. [CrossRef Medline](#)
- Kish SJ, Shannak K, Hornykiewicz O (1988) Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson’s disease: pathophysiologic and clinical implications. *N Engl J Med* 318:876–880. [CrossRef Medline](#)
- Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006) Midbrain dopamine neurons encode decisions for future action. *Nat Neurosci* 9:1057–1063. [CrossRef Medline](#)
- O’Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454. [CrossRef Medline](#)
- Piray P, Zeighami Y, Bahrami F, Eissa AM, Hewedi DH, Moustafa AA (2014) Impulse control disorders in Parkinson’s disease are associated with dysfunction in stimulus valuation but not action valuation. *J Neurosci* 34:7814–7824. [CrossRef Medline](#)
- Ramayya AG, Misra A, Baltuch GH, Kahana MJ (2014) Microstimulation of the human substantia nigra alters reinforcement learning. *J Neurosci* 34:6887–6895. [CrossRef Medline](#)
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599. [CrossRef Medline](#)
- Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 16:966–973. [CrossRef Medline](#)
- Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N (2012) Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74:858–873. [CrossRef Medline](#)
- Zaghloul KA, Blanco JA, Weidemann CT, McGill K, Jaggi JL, Baltuch GH, Kahana MJ (2009) Human substantia nigra neurons encode unexpected financial rewards. *Science* 323:1496–1499. [CrossRef Medline](#)