

Journal Club

Editor's Note: These short reviews of recent *JNeurosci* articles, written exclusively by students or postdoctoral fellows, summarize the important findings of the paper and provide additional insight and commentary. If the authors of the highlighted article have written a response to the Journal Club, the response can be found by viewing the Journal Club at www.jneurosci.org. For more information on the format, review process, and purpose of Journal Club articles, please see <https://www.jneurosci.org/content/jneurosci-journal-club>.

What Can Saccades Reveal about the Link between Learning and Motivation?

 **Huw Jarvis**

Turner Institute for Brain and Mental Health, Monash University, Victoria 3800, Australia
Review of Sedaghat-Nejad et al.

Making a choice involves weighing up the value of each outcome against the costs required to achieve it, such as time and effort. Through this process we decide not only what to do but how to do it. For example, actions with higher value tend to be executed more quickly, including reaching movements (Summerside et al., 2018) and visual saccades (Milstein and Dorris, 2007). Separate work has shown that value is closely linked to the motivation to act, such that the more we value something, the harder we will work to obtain it (Chong et al., 2015). These findings are intuitive when viewed through the lens of reinforcement learning: responding with greater vigor helps us to maximize the amount of reward acquired (Sutton and Barto, 1998).

The term *vigor* refers to expending energy to overcome time and effort costs during motivated behavior. Growing evidence suggests that dopaminergic reward signals underpin such computations. Dopamine agonists increase the sensitivity of response times to changes in reward magnitude (Beierholm et al., 2013) and restore

willingness to exert effort for reward in patients with Parkinson's disease (Chong et al., 2015). Similarly, saccades to visual stimuli are faster when the magnitude of anticipated reward is higher, but only when dopamine signaling is intact (Nakamura and Hikosaka, 2006).

Although much of this work has focused on anticipation or acquisition of reward, less is known about how vigor responds to the difference between these quantities, reward prediction error (RPE), which is conveyed by rapid changes in dopamine firing rates (Schultz et al., 1997). If dopamine indeed modulates vigor, acquiring a reward larger than anticipated [positive prediction error (+RPE)] should increase vigor, and acquiring one smaller than anticipated [negative prediction error (−RPE)] should decrease vigor. In other words, rather than the size of the reward anticipated or acquired, movements should be sensitive to the direction and magnitude of the RPE. This was the prediction made by Sedaghat-Nejad et al. (2019) in a recent paper published in the *Journal of Neuroscience*.

RPEs are typically computed at the end of an action when the outcome becomes known. This makes it difficult to test their effect on vigor, since the action has already been completed by the time the RPE signal is conveyed. Sedaghat-Nejad et al. (2019) overcame this by designing a double-saccade paradigm in humans to elicit a RPE in the milliseconds before the

secondary saccade. Relying on evidence that it is more rewarding to view faces than other images (O'Doherty et al., 2003; Yoon et al., 2018), the researchers induced visual saccades to images of an intact face (face image) or a scrambled face (noise image). After onset of the primary saccade on each trial, the first image was removed probabilistically and a second image appeared on the screen nearby, inducing a secondary saccade.

RPEs occurred because there was a chance the second image would be different from the first, which meant there was a discrepancy between the reward value predicted on perceiving the first image and the actual reward obtained by gazing at the second image. For example, if the first image were a face, the anticipated reward would be slightly less than its actual value because of the possibility it would change to a noise image. If the second image turned out to be a face after all, the result would be a small +RPE. As such, there were four trial types with different RPEs: noise-face (large +RPE), face-face (small +RPE), noise-noise (small −RPE), and face-noise (large −RPE).

Vigor of the secondary saccade was defined as the time from completion of the primary saccade to arrival at the second image. The authors examined reaction time and peak velocity as distinct components of vigor. On both measures, the secondary saccade varied significantly in the predicted direction. The highest vigor

Received July 6, 2019; revised Sept. 18, 2019; accepted Sept. 24, 2019.

This work was supported by a PhD Scholarship awarded by the Rebecca L. Cooper Medical Research Foundation and an Australian Government Research Training Program Scholarship.

The author declares no competing financial interests.

Correspondence should be addressed to Huw Jarvis at huw.jarvis@monash.edu.

<https://doi.org/10.1523/JNEUROSCI.1598-19.2019>

Copyright © 2019 the authors

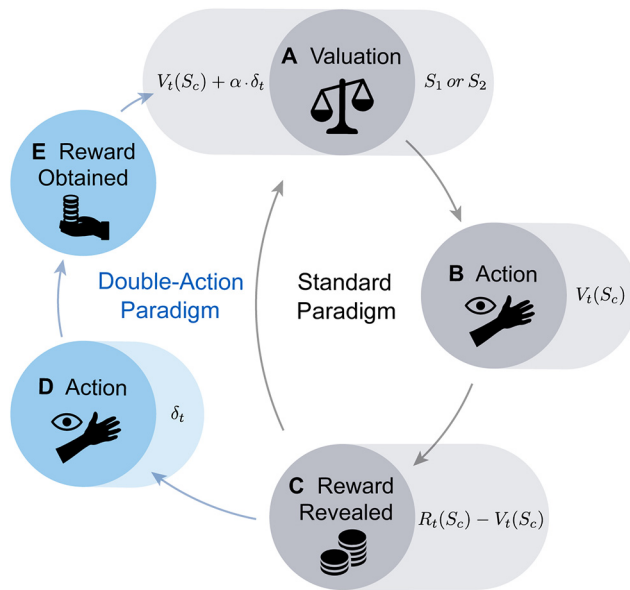


Figure 1. ‘Double-action’ paradigms may offer a way to characterize vigor modulation during reinforcement learning. **A**, The participant chooses a stimulus [S_1 or S_2]. **B**, The vigor of their response is modulated by the value of the chosen stimulus [$V_i(S_c)$]. **C**, The reward is revealed. The participant computes a reward prediction error by comparing expected with actual reward magnitude [$\delta_t = R_t(S_c) - V_i(S_c)$]. In a standard design, reward is obtained at this point. In a double-action design, a second action is required to obtain the reward. **D**, The vigor of the second action is modulated by the direction and magnitude of the reward prediction error [δ_t]. **E**, The reward is obtained at the end of the trial. An important question is whether vigor is modulated by the reward prediction error [δ_t] or the resulting update in expected value [$V_i(S_c) + \alpha \cdot \delta_t$].

(i.e., shortest reaction time and highest peak velocity) followed the largest +RPE, and the lowest vigor followed the largest –RPE. Crucially, reaction time was also significantly shorter on noise-face compared with face-face trials (i.e., large vs small +RPEs) and on noise-noise compared with face-noise trials (small vs large –RPEs), showing that vigor was modulated by the magnitude of the RPE, not just the value of the second image.

This finding suggests that rapid changes in dopamine firing rates associated with RPEs may play a role in motivating action. The classical account is that while dopamine signals underpin both learning and motivation, these operate over different timescales (Schultz, 2007). Namely, learning is driven by phasic RPE signals (Schultz et al., 1997) and motivation is linked to slower dopamine release in the striatum (Niv et al., 2007; Howe et al., 2013). In contrast, the current finding shows that saccade vigor in humans is sensitive to RPE signals on a subsecond timescale (Sedaghat-Nejad et al., 2019). This is consistent with emerging evidence from rodent studies that phasic bursts of dopamine also play a role in invigorating behavior (Howe and Dombeck, 2016; da Silva et al., 2018). However, an important question that Sedaghat-Nejad et al. (2019) did not discuss is why RPEs should modulate vigor.

One possibility is that RPEs are closely related to changes in average reward rate. Previous work showed that vigor is modulated according to the average reward rate of the environment, which is conveyed by slow changes in striatal dopamine activity (Niv et al., 2007). When reward rates increase, responses become faster to maximize the amount of reward acquired. Recent evidence suggests that fast changes in striatal dopamine may modulate vigor by the same logic (Hamid et al., 2016). Hamid et al. (2016) found that in addition to reinforcing rewarded choices, striatal dopamine fluctuations immediately altered the response vigor of rats during choice behavior. In addition to more gradual changes in reward rate and reward proximity, dopamine levels tracked rapid updates in expected value, which were driven by RPEs. Movement vigor responded immediately to these updates in value.

Although the double-saccade experiment of Sedaghat-Nejad et al. (2019) did not explicitly encourage learning, the link between RPEs and value-updating is clearly demonstrated in a simple reinforcement learning model (Rescorla and Wagner, 1972; Sutton and Barto, 1998):

$$V_{t+1}(s) = V_t(s) + \alpha \cdot \delta_t$$

$$\delta_t = R_t(s) - V_t(s)$$

The model states that the expected value of a stimulus on the next trial [$V_{t+1}(s)$] will be updated according to the RPE on the current trial [δ_t]. The RPE is calculated as the difference between the reward acquired [$R_t(s)$] and the current expected value [$V_t(s)$]. The extent to which the RPE updates expected value is determined by the learning rate [α], which adjusts the magnitude of the change in expected value on each trial.

In this model, the expected value of a stimulus [$V_t(s)$] represents a cached average of the reward available from that stimulus. The RPE [δ_t] indicates how much that average might be updated on the next trial. In this sense, RPEs represent instantaneous updates to average reward rate. If one accepts that vigor should reflect average reward rate, it follows that vigor might vary according to the magnitude and direction of RPEs. In other words, the same signal that drives reward-based learning could also motivate behavior, as demonstrated by Sedaghat-Nejad et al. (2019).

The notion that the same dopamine signals can convey information about reward and motivation is supported by a recent study that used a Go/No-Go task in rats (Syed et al., 2016). The study found that rapid increases in nucleus accumbens dopamine levels were only associated with reward cues when an action was required, not when an action was suppressed. The cues were identical with respect to the magnitude and timing of rewards; the only difference was the requirement to act. Importantly, however, these dopamine signals related to reward anticipation rather than RPEs. In contrast, a different study recently dissociated RPE signals conveyed by midbrain dopamine bursts from motivation signals in striatum (Mohebi et al., 2019). In sum, the precise links between reward signals in learning and motivation remain unclear.

Future studies could contribute to this work by using the double-saccade experiment of Sedaghat-Nejad et al. (2019) to characterize vigor modulation in humans during reinforcement learning. For example, an important question is whether vigor responds more closely to RPEs or to resulting updates in expected value. To test this, similar ‘double-action’ paradigms could be based on the same approach: the reward outcome becomes known (e.g., revealed on screen) but a final action is required before it is obtained (e.g., reaction time test; Fig. 1).

In summary, the recent study by Sedaghat-Nejad et al. (2019) provides an

elegant paradigm to investigate dopamine dynamics behaviorally. The study demonstrates that saccade vigor is modulated by RPEs, consistent with recent rodent studies showing that phasic dopamine signals play a role in invigorating behavior (Howe and Dombek, 2016; da Silva et al., 2018). Future research could use a similar experiment to characterize vigor modulation in humans during reinforcement learning. This could make a valuable contribution to ongoing work in rodent studies attempting to disentangle the reward signals that underpin learning and motivation (Hamid et al., 2016; Mohebi et al., 2019).

References

- Beierholm U, Guitart-Masip M, Economides M, Chowdhury R, Düzél E, Dolan R, Dayan P (2013) Dopamine modulates reward-related vigor. *Neuropsychopharmacology* 38:1495–1503.
- Chong TT, Bonnelle V, Manohar S, Veromann KR, Muhammed K, Tofaris GK, Hu M, Husain M (2015) Dopamine enhances willingness to exert effort for reward in Parkinson's disease. *Cortex* 69:40–46.
- da Silva JA, Tecuapetla F, Paixão V, Costa RM (2018) Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* 554:244–248.
- Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT, Aragona BJ, Berke JD (2016) Mesolimbic dopamine signals the value of work. *Nat Neurosci* 19:117–126.
- Howe MW, Dombek DA (2016) Rapid signaling in distinct dopaminergic axons during locomotion and reward. *Nature* 535:505–510.
- Howe MW, PTierney PL, Sandberg SG, Phillips PE, Graybiel AM (2013) Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* 500:575–579.
- Milstein DM, Dorris MC (2007) The influence of expected value on saccadic preparation. *J Neurosci* 27:4810–4818.
- Mohebi A, Pettibone JR, Hamid AA, Wong JT, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD (2019) Dissociable dopamine dynamics for learning and motivation. *Nature* 570:65–70.
- Nakamura K, Hikosaka O (2006) Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *J Neurosci* 26:5360–5369.
- Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* 191:507–520.
- O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ (2003) Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41:147–155.
- Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations on 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 (2007) Multiple dopamine functions at different time courses. *Annu Rev Neurosci* 30:259–288.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
- Sedaghat-Nejad E, Herzfeld DJ, Shadmehr R (2019) Reward prediction error modulates saccade vigor. *J Neurosci* 39:5010–5017.
- Summerside EM, Shadmehr R, Ahmed AA (2018) Vigor of reaching movements: reward discounts the cost of effort. *J Neurophysiol* 119:2347–2357.
- Sutton RS, Barto AG (1998) *Reinforcement learning: an introduction*. Cambridge, MA: MIT.
- Syed EC, Grima LL, Magill PJ, Bogacz R, Brown P, Walton ME (2016) Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat Neurosci* 19:34–36.
- Yoon T, Geary RB, Ahmed AA, Shadmehr R (2018) Control of movement vigor and decision making during foraging. *Proc Natl Acad Sci U S A* 115:E10476–E10485.