Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

The biological and behavioral computations that influence dopamine responses

WR Stauffer

Phasic dopamine responses demonstrate remarkable simplicity; they code for the differences between received and predicted reward values. Yet this simplicity belies the subtle complexity of the psychological, computational, and contextual factors that influence this signal. Advances in behavioral paradigms and models, in monkeys and rodents, have demonstrated that phasic dopamine responses reflect numerous behavioral computations and factors including choice, subjective value, confidence, and context. The application of optogenetics has provided evidence that dopamine reward prediction error responses cause value learning. Furthermore, studies using advanced circuit tracing techniques have begun to uncover the biological network implementation of the reward learning algorithm. The purpose of this review is to summarize the recent advances in dopamine neurophysiology and synthesize an updated account of the behavioral function of dopamine signals.

Address
Department of Neurobiology, Systems Neuroscience Center, University of Pittsburgh, Pittsburgh, PA 15261, United States

Corresponding author: Stauffer, WR (wrs@pitt.edu)

Current Opinion in Neurobiology 2018, **49**:123–131

This review comes from a themed issue on **Neurobiology of behavior**

Edited by **Kay Tye** and **Naoshige Uchida**

https://doi.org/10.1016/j.conb.2018.02.005

Introduction

Reward prediction errors are arguably one of the oldest biological computations on Earth. The single cell bacteria that dominated life for over two billion years detected and responded to positive and negative differences, in time and space, in the concentrations of environmental substances [1]. Positive concentration changes evoke approach behavior in the form of movements towards the source, whereas increasing concentrations of harmful chemicals cause bacteria to avoid the source and 'tumble' away in random directions [2,3]. Much has changed in two billion (or so) years of evolution, but computation of unpredicted changes for better or worse remains critical to optimal behavioral function and is broadly employed in the brain. Phasic dopamine responses constitute the prime example of neuronal reward prediction error coding.

Dopamine neurons are predominantly located in the midbrain A8, A9, and A10 cell groups that correspond roughly to the Retrorubral Field (RRF), the *Substantia Nigra pars compacta* (SNc) and Ventral Tegmental Area (VTA), respectively [4]. These neurons receive synaptic input from over 30 different brain regions [5–11], and send the majority of their projections to basal ganglia and frontal cortex areas involved in motor control, learning, and cognitive function [11–14]. They respond to rewards and reward predicting cues with phasic bursts of action potentials that code for reward prediction errors, the differences between received and predicted rewards [15,16•]. Dopamine prediction error responses are an ideal mechanism to guide behaviors to harvest more and better rewards. Positive prediction error responses indicate that the preceding action should be repeated or invigorated, whereas negative prediction error responses indicate that the preceding behavior should be decreased or avoided [17].

Recent studies have shown that numerous behavioral computations, including value, choice, confidence, and contextual expectations are factored into the canonical reward prediction error (RPE) response in dopamine neurons [18••,19–21,22••,23]. Next generation technologies have been critical to understanding the behavioral functions of dopamine neurons [24••,25••,26••,27,28], their downstream effects [29,30], and how they compute RPEs [6••,9,31,32••,33,34••]. This non-comprehensive review endeavors to highlight the recent novel findings in dopamine physiology as it pertains to reward coding and its behavioral consequences.

Phasic dopamine signals are reinforcement learning signals

Prediction errors are the fundamental element of reinforcement learning models, including the Rescorla–Wagner [35] and temporal difference (TD) [17] models. Prediction errors are used to update (i.e., learn) the value of predictive stimuli. The prediction error in TD models provides a theoretical account for phasic dopamine activity [36,37]. The TD prediction error

(TDPE) is a difference between the predicted and actual value:

$$\text{TDPE} = V_{\text{actual}} - V_{\text{predicted}}$$

Thus, subtraction is the fundamental operation that guides value updating. To verify that subtraction governs the response of dopamine neurons, the activity of optogenetically identified mouse dopamine neurons was recorded during reward delivery. The delivered rewards were either (a) completely unpredicted — neither the magnitude nor timing was known, or (b) followed a cue (odor) that predicted the average magnitude and exact timing — the prediction was a constant and only the exact reward magnitude was unknown. The constant reward expectation generated by the predictive odor reduced, by an equivalent amount, the dopamine response to every reward magnitude [32••]. This result indicates that dopamine neurons perform subtraction of expected reward value from actual value, as opposed to using divisive operations that are more commonly observed in neural circuits [38,39]. Moreover, every recorded dopamine neuron used a similar subtractive algorithm [31]. These results confirm that, just like the prediction error signal that forms the core of reinforcement learning models, the magnitude of the phasic dopamine response is governed by subtraction.

More than two decades of research has provided strong correlational evidence that phasic dopamine responses constitute a reward learning signal (for a concise summary, see [16•], a more comprehensive review is provided in [15]). However, new techniques like optogenetics finally permit us to ask whether dopamine signals *cause* learning to occur. Prediction error responses have been simulated using optogenetic techniques in a variety of behavioral tasks in mice [27,28,40], rats [26••,41,42], and monkeys [25••]. In every species tested, phasic optogenetic stimulation or suppression of dopamine neurons has resulted in behavioral observations consistent with a critical role for dopamine neurons in reward learning.

A fundamental insight from animal learning theory is that rewards must be unpredicted, that is, they must generate reward prediction errors, for learning to occur [35]. The strongest evidence for the causal role of dopamine in learning comes from experimental manipulations in behavioral paradigms where prediction errors would not normally occur, and where no learning would normally happen. These paradigms reveal how introduction of phasic activations or suppressions of dopamine neurons affect learning. Effects of optical activation and suppression of dopamine neurons in rats have been tested during a blocking and an overestimation paradigm, respectively [24••,26••].

During blocking, formation of associative strength between a conditioned stimulus (CS) and an unconditioned stimulus (US) is 'blocked' by a secondary stimulus that fully predicts the US. Dopamine neurons do not respond to CSs that have been blocked [43]. Artificial phasic dopamine activations unblock the CS leading to increased conditioned responses (time spent in the reward port) and indicating learning of the unblocked CS–US association [26••]. Thus, optogenetic activation of phasic dopamine mimics the effects of positive prediction errors, and is sufficient to cause associative learning. During overexpectation, the compound presentation of two reward-predicting CSs generates heightened expectation — 'overexpectation' — that likely corresponds to both rewards being delivered. The negative prediction errors associated with delivery of only one reward leads to extinction of the original CS-reward associations [35,44–46]. In a modified overexpectation paradigm, the two rewards were actually delivered, fulfilling the heightened expectations, and this modification eliminated the extinction. Phasic optogenetic silencing of dopamine neurons reinstates extinction learning in the modified overexpectation task [24••]. Thus, optogenetic silencing of dopamine mimics the effects of negative prediction errors, and is sufficient to cause extinction learning. Together, these findings provide evidence that phasic dopamine activations and suppressions constitute bidirectional teaching signals that cause increases and decreases (respectively) in the associative strengths between rewards and their predictors.

In many situations, including the behavioral tasks reviewed so far, dopamine signals update predictions using a 'model-free'-like algorithm. That is, cue-outcome associations are updated according to direct experience of the cues and outcomes. In contrast, some outcomes can be used to update a model that contains multiple associations. Such 'model-based' learning can occur, for instance, during a reversal learning task (for a review of model-free vs model-based reinforcement learning, see [47]). Monkeys learned that one cue predicted reward while another cue predicted no reward, and on a randomly selected trial the reward contingencies reversed. Model-based learning can use the outcome of the first reversal trial to update the value of both stimuli, even with no direct experience of the other cue-outcome association. Dopamine responses reflected values updated according to this model-based rule [48]. This result suggests that the dopamine system is adapted to efficiently learn environmental reward contingencies whether they are experienced directly or merely inferred. Accordingly, this neuronal teaching signal can support multiple forms of reinforcement learning and likely updates value correlates throughout the brain.