Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
Curr Opin Neurobiol. Author manuscript; available in PMC 2018 August 16.

Published in final edited form as:
Curr Opin Neurobiol. 2018 April ; 49: 123–131. doi:10.1016/j.conb.2018.02.005.

The biological and behavioral computations that influence dopamine responses

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

Abstract

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.

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

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Corresponding author: Stauffer, WR (wrs@pitt.edu).
Conflict of interest statement
Nothing declared.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 diference 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 obser-vations 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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.

Dopamine responses reflect behavioral computations associated with value

Most rewards do not possess a common physical scale for direct comparison, and they often enter awareness via biophysically distinct pathways. Food, drink, money, and social interaction are but a few examples of the larger category of objects, events, or thoughts that we readily recognize as rewards [15]. Despite the heterogeneity of reward features, individuals quickly appreciate reward value and readily exchange one reward for another. For example, individuals readily handover money in exchange for ice cream. This behavior implies that coding of reward value is not critically dependent of the sensory properties of rewards. When monkeys make choices between different reward types, dopamine responses are larger to more preferred rewards, compared to less preferred rewards. Importantly, when choices indicate indifference between two rewards, the dopamine responses to those rewards are indistinguishable (Figure 1a) [20]. Similarly, when rats have been fed to satiety on one reward, their choices indicate that value of the overfed reward is decreased, and the dopamine response to the overfed reward is also decreased [49]. These patterns of activity suggest that dopamine responses reflect the subjective value of rewards.

To demonstrate the functional relationship between subjective value and dopamine activity, subjective value was measured as a function of physical value in monkeys making choices between risky and safe outcomes. There is a mathematical relationship between risk attitudes, whether the monkey is risk seeking or risk avoiding, and the curvatures of the resulting value functions [50]. Choices between risky rewards revealed an ‘S’-shaped subjective value (utility) function that reflected risk seeking for small rewards and risk avoiding for large rewards (Figure 1b, red line) [23,51]. The magnitudes of dopamine responses to unpredicted rewards were correlated with the shape of the measured utility functions (Figure 1b) [23]. During behavioral choices, dopamine responses scaled with the value of the chosen options [21,52]. Moreover, dopamine responses were larger on trials when the monkey indicated the correct choice, compared to when it was mistaken (Figure 1c) [19]. These results demonstrate that dopamine responses integrate moment-by-moment behavioral information with reinforcement learning to code the same dynamic value information that is used to make decisions.

Value may be gained from physical rewards, but also may be derived from the internal evaluation of performance. A novel study examined dopamine activity during bird song learning. As juvenile birds learned to sing, distorted audio feedback (DAF) was provided at unpredictable times. Dopamine neurons paused their firing when they heard the DAF, as though they were responding to a negative prediction error (Figure 1d). The response was contingent upon the bird singing; dopamine neurons were unaffected by DAF when they were not singing [18••]. This result demonstrates that dopamine neurons are active during

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

performance monitoring, but it remains unclear whether this response reflected the performance error itself or whether it reflected the value of that error. Developing behavioral technologies to measure the value of good performance is key to understanding how reward and motor systems interact to motivate motor learning.

Value has many sources, including long term reinforcement learning history, context, and trial-by-trial behavioral factors. Overall, these recent studies demonstrate that dopamine prediction error responses reflect these many sources of value, and provide deeper insights into the nature of biological reinforcement learning.

Biological implementation of reward prediction error computations

Dopamine neurons receive input from more than thirty brain areas, including the lateral hypothalamus, subthalamic nucleus, the pedunculopontine nucleus, the lateral habenula, the striatum, and the dorsal raphe nucleus [6••,7,11,53,54]. It is of considerable interest how dopamine neurons integrate information from these diverse brain regions to compute RPE.

Electrophysiological analysis of more than 200 neurons mono-synaptically connected to VTA dopamine neurons revealed that most input neurons coded for some computational components of RPEs, such as responses to unpredicted rewards or reward expectation. These different component responses were not localized to specific input nuclei, but rather were distributed across all the sampled nuclei [6••]. Thus, dopamine neurons receive distributed inputs from many structures in the brain that code for parts of the RPE, and appear to integrate these distributed inputs to compute RPE responses.

A critical factor determining the dopamine response is the cell type-identity of their inputs. For example, the lateral hypothalamus (LH) is a major source of input to the midbrain. Optogenetic activations of GABAergic projections from the LH to the VTA cause phasic dopamine release in the striatum and promote reward seeking behaviors [9,33]. By contrast, phasic activations of LH glutamatergic projections to the midbrain result in avoidance behaviors [9]. The appetitive and aversive effects of LH inputs seem to operate through di-synaptic mechanisms involving local GABA neurons in the vicinity of the VTA [9]. However, the exact identity and location of GABA neurons mediating the disinhibition that could lead to dopamine activations remains unclear. Although LH-GABA neurons synapse onto GABAergic neurons in the VTA [9], previous research has failed to find consistent, phasic inhibition of VTA-GABA neurons at the precise moments when such inhibitions could mediate dopamine activations [55].

Similar to the input from the LH, the different cell types in the dorsal raphe nucleus (DRN) differentially contribute to appetitive and aversive dopamine-mediated behaviors. Both serotonergic and glutamatergic projections from the DRN appear to provide appetitive information to dopamine neurons [53–56], whereas DRN GABA neuron activity is correlated with aversive responses [56].

Glutamatergic projection neurons in the lateral habenula (LHb) are a major source of aversive information to dopamine neurons. Activation of LHb inhibits the majority of dopamine neurons via di-synaptic connections with GABAergic neurons in the rostromedial

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

tegmental nucleus (RMTg) [57–59]. Activation of this pathway causes conditioned place aversion [10,60], whereas lesioning this pathway disrupts the normal inhibitory responses to unpredicted reward omission [8••]. Thus, the majority of LHb activation is related to negative prediction error responses (pauses) in dopamine neurons. Studies in the mouse have recently discovered a group of dopamine neurons that receive mono-synaptic, excitatory drive from the LHb [10,61]. Notably, this group of dopamine neurons has different electrophysiological properties than dopamine neurons that have been identified with apomorphine [62–65] and optogenetics [31,32••,55]. They emit action potentials at much higher rates than classical dopamine neurons, and do not express somatodendritic dopamine D2 receptors [61]. As such, they are likely insensitive to apomorphine. Because of the distinct electrophysiological properties and putative apomorphine insensitivity, these neurons are not likely to be sampled by studies that use traditional waveform identification techniques [62–65]. Therefore, it remains to be seen what the behavioral function of these neurons is, and indeed whether they code for prediction errors.

Diversity of dopamine responses

The majority of dopamine neurons are activated by reward [31,55,66]. However, dopamine neurons do not respond solely to rewards, nor do all dopamine neurons respond identically. In fact, numerous other stimuli elicit dopamine responses, including noxious, aversive, and physically salient stimuli [64,67–73], novelty [21,73], and large movements [74,75]. Likewise, the well-known role of dopamine neuron loss in movement disorders and Parkinson's disease clearly implicates these neurons in movement, albeit indirectly. These observations have generated intense interest into the functional heterogeneity of the dopamine population.

Phasic dopamine responses display complex temporal dynamics that may reflect different variables and even have multiple behavioral functions [19,21,70,71,76]. Dopamine neurons generally encode RPEs at latencies between 150 and 250 ms following reward; the longer latencies are observed when stimuli are hard to distinguish or the experiment highly dynamic [19,21,77]. On the other hand, many non-reward activations of dopamine neurons occur very early in the response, within 50–200 ms of the behavioral event. Aversive air-puffs evoke short-latency activations and inhibitions in dorsolateral and ventromedial dopamine neurons, respectively [72]. Similarly, aversive electrical shocks activate dopamine neurons projecting to the dorsolateral striatum (DLS), whereas the same stimuli inhibit dopamine neurons projecting to the dorsomedial striatum (DMS) [68]. The Ca^{2+} signal used to detect the electric shock activations and inhibitions has a slower time-course than action potential signals [78]. Nevertheless, in dopamine neurons that project to the DLS, activations driven by electric shock appear faster and shorter than activations driven by reward [68]. Thus, dopamine responses to aversive stimuli are heterogeneous, but short-latency non-reward responses appear to arise earlier than RPE coding.

Context sensitivity also plays an important role in dopamine response heterogeneity. Context can be defined by numerous factors including the overall reward availability, task dynamics, and the physical nature of stimuli and rewards. In particular, the overall reward availability strongly modulates dopamine responses to rewards, reward predictors, and nonrewards. For

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

example, when the overall availability of rewards was low (i.e., rewards were only delivered on a small fraction of trials) dopamine neurons responded to aversive stimuli with inhibitions. However, when the overall availability of rewards was high, the very same dopamine neuron respond to aversive stimuli with short-latency activations, rather than inhibitions [34••]. Neutral cues were similarly influenced by the amount of reward delivered in a specific context; greater overall reward availability resulted in greater responses to neutral cues [79]. Because of the influence of overall reward availability, the short-latency activations observed in these (and other) studies are likely instances of pseudo-conditioning, a process of generalization between USs, rather than true responses to nonrewarding stimuli [80]. In a similar fashion, sensory stimulus (CS) generalization has a major impact on the number of dopamine neurons that respond to aversive-predicting CS. When visual cues are used to predict both rewarding and aversive outcomes, more than half of dopamine neurons can respond to the aversive visual cue [66,72]. This is in contrast to when an auditory cue is used to predict a reward and a visual cue is used to predict an aversive air puff, only 15% of dopamine neurons respond to the aversive visual cue [66]. Thus, less stimulus generalization translates into fewer dopamine neurons responding to aversive events. These critical studies highlight the importance of considering behavioral and contextual factors, in addition to the underlying circuits, when designing behavioral tasks and interpreting the motivational implications of dopamine activity.

Conclusions

The value assigned to rewards is a highly dynamic quantity influenced by numerous factors. Dopamine responses code for subjective reward value (utility) [20,21,23] and reflect the numerous behavioral factors that influence value, including choice [21,81], confidence [19], context [34••,79], and satiety [49]. Optogenetic stimulation and suppression of dopamine neurons demonstrates that these signals cause value learning [24••,26••], which likely updates action values in the striatum and elsewhere [34••,82–84]. Beyond learning, dopamine neurons have many behavioral functions including roles in movement and motivation. Not discussed here, but at the cutting edge of dopamine investigations, are studies deciphering the precise behavioral roles of prediction error coding and dopamine release at the interface between motivation and movement [29,30].

Our current understanding of dopamine neurons has been greatly facilitated by recent technological developments, including optogenetics and advanced circuit tracing techniques. Optogenetic technologies have been used to unambiguously identify dopamine neurons and test their behavioral function [6••,22••,31,32••,34••,55]. These studies have confirmed the major role that dopamine neurons have in reward coding [55], and detailed the algorithm used by the dopamine population [31,32••]. Optogenetics stimulation has been used to test and confirm the hypothesis that phasic dopamine activations and suppressions constitute a bi-directional teaching signal for value learning [24••,26••]. Optogenetic stimulation of monkey dopamine neurons biases the choices of the animal to the stimulation-reinforced options, and translates these technological capabilities into a species with greater anatomical and functional homology with humans [25••].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In the physical dopamine circuit, perhaps more than anywhere else in the brain, we are starting to understand the Marr-level III implementation of the reward prediction error algorithm [85]. Recent studies have mapped the anatomical and functional inputs of dopamine neurons [6••,8••,9,10,33,34••,61]. These results demonstrate how different cell types and the micro-circuits they form are critical to understand how dopamine responses are shaped. An important question for future studies to address is, 'what level of circuit detail is relevant to behavioral function?' Advancing technological capabilities are revealing ever more complex circuit maps with ever finer details, but it is critical that these findings are interpreted in light of well-founded behavioral theories and experiments [86]. Nevertheless, these developments promise to provide deeper insights into the behavioral functions and information processing capacities of this critical neural system.

The next step, I believe, is to gain a broader and clearer appreciation of the nature of reward predictions. To say that dopamine neurons code for reward prediction errors is to imply subtraction is taking place. The operation of subtraction includes three terms, the minuend (the number being subtracted from), the subtrahend (the number being subtracted), and the resulting difference. Recent studies have confirmed that dopamine prediction error responses truly represent differences [31,32••]. Therefore, to fully understand the dopamine signal, we need to understand how dopamine neurons code the minuend (reward) and the subtrahend (prediction). Regarding the former, significant progress has been made in understanding how the brain codes for rewards. Signals related to subjective value have been recorded in multiple brain areas. Dopamine neurons reflect multiple attributes of reward, including reward magnitude [87], probability [88,89], and delay [90]. They integrate these attributes and code for a highly specified form of subjective reward value, economic utility, that places the value of different rewards on a common scale for easy comparison [20,23]. Well-defined and easily measured utility functions, therefore, provide a rigorous account of the dopaminergic minuend (Figure 1b). However, we know less about the dopaminergic subtrahend: reward predictions. The classical model for dopamine activity, the TD model, predicts the time-discounted expected value of future rewards [36]. Although this quantity is surely factored into dopamine signals, the results reviewed here demonstrate that this model provides an inadequate description of dopamine activity. Inference about hidden states of the world as well as factors related to decision confidence are incorporated into dopamine responses [19,22••]; both of these factors are well beyond simple first-order reward statistics. These results indicate that the reward predictions made by dopamine neurons, and by implication the brain, are far richer than was previously thought. Fortunately, the well-defined nature of the dopamine signal provides an excellent substrate to learn about the shape and character of neuronal predictions.

Acknowledgements

This work was funded by the University of Pittsburgh Brain Institute and by the National Institutes of Health through the NIH Director's New Innovator Award 1DP2MH113095.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

• of special interest

•• of outstanding interest

1. Segall JE , Block SM , Berg HC : Temporal comparisons in bacterial chemotaxis. Proc Natl Acad Sci USA 1986, 83:8987–8991. [PubMed: 3024160]

2. Macnab RM , Koshland DE : The gradient-sensing mechanism in bacterial chemotaxis. Proc Natl Acad Sci U S A 1972, 69:2509–2512. [PubMed: 4560688]

3. Adler J : Chemotaxis in bacteria. Annu Rev Biochem 1975, 44:341–356. [PubMed: 1094913]

4. Dahlstroem A , Fuxe K : Evidence for the existence of monoamine-containing neurons in the central nervous system. I. Demonstration of monoamines in the cell bodies of brain stem neurons. Acta Physiol Scand Suppl 1964, Suppl. 232:231–255.

5. Dautan D , Souza AS , Huerta-Ocampo I , Valencia M , Assous M , Witten IB , Deisseroth K , Tepper JM , Bolam JP , Gerdjikov TV et al.:Segregated cholinergic transmission modulates dopamine neurons integrated in distinct functional circuits. Nat Neurosci 2016, 19:1025–1033. [PubMed: 27348215]

6••. Tian J , Huang R , Cohen JY , Osakada F , Kobak D , Machens CK , Callaway EM , Uchida N , Watabe-Uchida M : Distributed and mixed information in monosynaptic inputs to dopamine neurons. Neuron 2016, 91:1374–1389. [PubMed: 27618675] A modified rabies virus infected neurons mono-synaptically connected to dopamine neurons (‘input neurons’) and caused the input neurons to express ChR2. The authors used optical stimulation to identify input neurons during behavior. Information that dopamine neurons use to compute RPE was distributed across many brain areas. Dopamine responses could be reconstructed from the input.

7. Watabe-Uchida M , Zhu L , Ogawa SK , Vamanrao A , Uchida N :Whole-brain mapping of direct inputs to midbrain dopamine neurons. Neuron 2012, 74:858–873. [PubMed: 22681690]

8••. Tian J , Uchida N : Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. Neuron 2015, 87:1304–1316. [PubMed: 26365765] Habenula lesions in mice disrupted the pause in action potentials that dopamine neurons exhibit to unpredicted reward omissions. The same lesions left the rest of the RPE response, including pauses to aversive stimuli, intact. This result provides evidence that dopamine neurons compute the RPE using distributed inputs.

9. Nieh EH , Vander Weele CM , Matthews GA , Presbrey KN , Wichmann R , Leppla CA , Izadmehr EM , Tye KM : Inhibitory input from the lateral hypothalamus to the ventral tegmental area disinhibits dopamine neurons and promotes behavioral activation. Neuron 2016, 90:1286–1298. [PubMed: 27238864]

10. Lammel S , Lim BK , Ran C , Huang KW , Betley MJ , Tye KM , Deisseroth K , Malenka RC : Input-specific control of reward and aversion in the ventral tegmental area. Nature 2012, 491:212–217. [PubMed: 23064228]

11. Beier KT , Steinberg EE , DeLoach KE , Xie S , Miyamichi K , Schwarz L , Gao XJ , Kremer EJ , Malenka RC , Luo L : Circuit architecture of VTA dopamine neurons revealed by systematic input-output mapping. Cell 2015, 162:622–634. [PubMed: 26232228]

12. Lynd-Balta E , Haber SN : The organization of midbrain projections to the ventral striatum in the primate. Neuroscience 1994, 59:609–623. [PubMed: 7516505]

13. Lynd-Balta E , Haber SN : The organization of midbrain projections to the striatum in the primate: sensorimotor-related striatum versus ventral striatum. Neuroscience 1994, 59:625–640. [PubMed: 7516506]

14. Lynd-Balta E , Haber SN : Primate striatonigral projections: a comparison of the sensorimotor-related striatum and the ventral striatum. J Comp Neurol 1994, 345:562–578. [PubMed: 7962700]

15. Schultz W : Neuronal reward and decision signals: from theories to data. Physiol Rev 2015, 95:853–951. [PubMed: 26109341]

16•. Schultz W : Reward prediction error. Curr Biol 2017, 27:R369–R371. [PubMed: 28535383] A simple and straightforward primer on reward prediction errors.

17. Sutton R , Barto A : Reinforcement Learning: An Introduction. MIT Press; 1998.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

18••. Gadagkar V , Puzerey PA , Chen R , Baird-Daniel E , Farhang AR , Goldberg JH : Dopamine neurons encode performance error in singing birds. Science 2016, 354:1278–1282. [PubMed: 27940871] During song learning, finch dopamine neurons are inhibited by simulated mistakes in their singing. This result demonstrates that dopamine neurons are sensitive to performance, and could provide a link between dopamine activity and motor learning.

19. Lak A , Nomoto K , Keramati M , Sakagami M , Kepecs A : Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. Curr Biol 2017, 27:821–832. [PubMed: 28285994]

20. Lak A , Stauffer WR , Schultz W : Dopamine prediction error responses integrate subjective value from different reward dimensions. Proc Natl Acad Sci USA 2014, 111:2343–2348. [PubMed: 24453218]

21. Lak A , Stauffer WR , Schultz W : Dopamine neurons learn relative chosen value from probabilistic rewards. Elife 2016:5.

22••. Starkweather CK , Babayan BM , Uchida N , Gershman SJ : Dopamine reward prediction errors reflect hidden-state inference across time. Nat Neurosci 2017, 20:581–589. [PubMed: 28263301] The exact timing of reward delivery was not predicted, and in 10% of trials, it was not delivered at all. Thus, as the trials progressed, there was greater subjective evidence that reward would never be delivered, and greater surprise inthe 90% of trials when it was. This temporally dynamic inference about whether the trial was rewarded or not was reflected in the dopamine response. This result demonstrates that dopamine RPE responses incorporate highly dynamic information, and provide deeper insights into the nature of the biological reinforcement learning algorithm.

23. Stauffer WR , Lak A , Schultz W : Dopamine reward prediction error responses reflect marginal utility. Curr Biol 2014,24:2491–2500. [PubMed: 25283778]

24••. Chang CY , Esber GR , Marrero-Garcia Y , Yau HJ , Bonci A , Schoenbaum G : Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat Neurosci 2016, 19:111–116. [PubMed: 26642092] Using an ‘overexpectation’ paradigm, the authors show that phasic dopamine pauses — negative prediction error responses — are sufficient to cause extinction.

25••. Stauffer WR , Lak A , Yang A , Borel M , Paulsen O , Boyden ES , Schultz W: Dopamine neuron-specific optogenetic stimulation in rhesus macaques. Cell 2016, 166:1564–1571 e1566. [PubMed: 27610576] A novel combination of viral vectors was used to achieve cell type-specific channelrhodopsin expression in Rhesus macaque dopamine neurons. Dopamine neurons responded more strongly to stimuli that predicted optical stimulation, than to stimuli that did not, and the monkeys’ choice behavior reflected this; they chose the options associated with stimulation with greater frequency.

26••. Steinberg EE , Keiflin R , Boivin JR , Witten IB , Deisseroth K , Janak PH : A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 2013, 16:966–973. [PubMed: 23708143] Using a ‘blocking’ paradigm, the authors show that phasic dopamine activations — positive prediction error responses — are sufficient to cause learning.

27. Tan KR , Yvon C , Turiault M , Mirzabekov JJ , Doehner J , Labouebe G , Deisseroth K , Tye KM , Luscher C : GABA neurons of the VTA drive conditioned place aversion. Neuron 2012, 73:1173–1183. [PubMed: 22445344]

28. van Zessen R , Phillips JL , Budygin EA , Stuber GD : Activation of VTA GABA neurons disrupts reward consumption. Neuron 2012, 73:1184–1194. [PubMed: 22445345]

29. Hamid AA , Pettibone JR , Mabrouk OS , Hetrick VL , Schmidt R , Vander Weele CM , Kennedy RT , Aragona BJ , Berke JD : Mesolimbic dopamine signals the value of work. Nat Neurosci 2016, 19:117–126. [PubMed: 26595651]

30. Parker NF , Cameron CM , Taliaferro JP , Lee J , Choi JY , Davidson TJ , Daw ND , Witten IB : Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat Neurosci 2016, 19:845–854. [PubMed: 27110917]

31. Eshel N , Tian J , Bukwich M , Uchida N : Dopamine neurons share common response function for reward prediction error. Nat Neurosci 2016, 19:479–486. [PubMed: 26854803]

32••. Eshel N , Bukwich M , Rao V , Hemmelder V , Tian J , Uchida N : Arithmetic and local circuitry underlying dopamine prediction errors. Nature 2015, 525:243–246. [PubMed: 26322583] Reward prediction errors are defined as received minus predicted reward. Therefore, subtraction is the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

fundamental operation guiding the prediction error algorithm. Recordings from optogenetically identified dopamine neurons revealed that the reward predictions reduced dopamine reward responses, to range of different reward sizes, by a constant amount. This finding confirms subtraction as the algorithmic principle governing phasic dopamine responses.

33. Sharpe MJ , Marchant NJ , Whitaker LR , Richie CT , Zhang YJ , Campbell EJ , Koivula PP , Necarsulmer JC , Mejias-Aponte C , Morales M et al.: Lateral hypothalamic GABAergic neurons encode reward predictions that are relayed to the ventral tegmental area to regulate learning. Curr Biol 2017, 27:2089–2100 e2085. [PubMed: 28690111]

34••. Matsumoto H , Tian J , Uchida N , Watabe-Uchida M : Midbrain dopamine neurons signal aversion in a reward-context-dependent manner. Elife 2016:5.Dopamine responses to aversive stimuli were dependent on reward context. In behavioral contexts where reward was rarely delivered, dopamine neurons were mostly inhibited by air-puff responses. In behavioral contexts where reward was delivered often, many dopamine neurons responded with a short-latency activation to air-puff. This result demonstrates that dopamine responses to aversive stimuli are sensitive to reward context, and highlight the importance of considering behavioral context when interpreting dopamine activations.

35. Rescorla RA , Wagner AR : A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non reinforcement In Classical Conditioning II: Current Research and Theory. Edited by Black AH , Prokasy WF . Appleton-Century-Crofts; 1972:64–99.

36. Schultz W , Dayan P , Montague PR : A neural substrate of prediction and reward. Science 1997, 275:1593–1599. [PubMed: 9054347]

37. Montague PR , Dayan P , Sejnowski TJ : A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 1996, 16:1936–1947. [PubMed: 8774460]

38. Louie K , Grattan LE , Glimcher PW : Reward value-based gain control: divisive normalization in parietal cortex. J Neurosci 2011, 31:10627–10639. [PubMed: 21775606]

39. Heeger DJ : Normalization of cell responses in cat striate cortex. Vis Neurosci 1992, 9:181–197. [PubMed: 1504027]

40. Tsai HC , Zhang F , Adamantidis A , Stuber GD , Bonci A , de Lecea L , Deisseroth K : Phasic firing in dopaminergic neurons is sufficient for behavior conditioning. Science 2009, 324: 1080–1084. [PubMed: 19389999]

41. Sharpe MJ , Chang CY , Liu MA , Batchelor HM , Mueller LE , Jones JL , Niv Y , Schoenbaum G : Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nat Neurosci 2017, 20:735–742. [PubMed: 28368385]

42. Witten IB , Steinberg EE , Lee SY , Davidson TJ , Zalocusky KA , Brodsky M , Yizhar O , Cho SL , Gong S , Ramakrishnan C et al.: Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. Neuron 2011, 72:721–733. [PubMed: 22153370]

43. Waelti P , Dickinson A , Schultz W : Dopamine responses comply with basic assumptions of formal learning theory. Nature 2001, 412:43–48. [PubMed: 11452299]

44. Lattal KM , Nakajima S : Overexpectation in appetitive Pavlovian and instrumental conditioning. Anim Learn Behav 1998, 26:351–360.

45. Rescorla RA : Spontaneous recovery from overexpectation. Learn Behav 2006, 34:13–20. [PubMed: 16786880]

46. Rescorla RA : Renewal after overexpectation. Learn Behav 2007, 35:19–26. [PubMed: 17557388]

47. Doll BB , Simon DA , Daw ND : The ubiquity of model-based reinforcement learning. Curr Opin Neurobiol 2012, 22:1075–1081. [PubMed: 22959354]

48. Bromberg-Martin ES , Matsumoto M , Hong S , Hikosaka O : A pallidus-habenula-dopamine pathway signals inferred stimulus values. J Neurophysiol 2010, 104:1068–1076. [PubMed: 20538770]

49. Papageorgiou GK , Baudonnat M , Cucca F , Walton ME : Mesolimbic dopamine encodes prediction errors in a state-dependent manner. Cell Rep 2016, 15:221–228. [PubMed: 27050518]

50. von Neumann J , Morgenstern O , Kuhn HW , Rubinstein A : Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition). Princeton University Press; 1944.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

51. Genest W , Stauffer WR , Schultz W : Utility functions predict variance and skewness risk preferences in monkeys. Proc Natl Acad Sci U S A 2016, 113:8402–8407. [PubMed: 27402743]
52. Morris G , Nevet A , Arkadir D , Vaadia E , Bergman H : Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 2006, 9:1057–1063. [PubMed: 16862149]
53. McDevitt RA , Tiran-Cappello A , Shen H , Balderas I , Britt JP , Marino RAM , Chung SL , Richie CT , Harvey BK , Bonci A : Serotonergic versus nonserotonergic dorsal raphe projection neurons: differential participation in reward circuitry. Cell Rep 2014, 8:1857–1869. [PubMed: 25242321]
54. Liu Z , Zhou J , Li Y , Hu F , Lu Y , Ma M , Feng Q , Zhang J-e , Wang D , Zeng J et al. : Dorsal raphe neurons signal reward through 5-HT and glutamate. Neuron 2014, 81:1360–1374. [PubMed: 24656254]
55. Cohen JY , Haesler S , Vong L , Lowell BB , Uchida N : Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 2012, 482:85–88. [PubMed: 22258508]
56. Li Y , Zhong W , Wang D , Feng Q , Liu Z , Zhou J , Jia C , Hu F , Zeng J , Guo Q et al.: Serotonin neurons in the dorsal raphe nucleus encode reward signals. Nat Commun 2016, 7:10503. [PubMed: 26818705]
57. Matsumoto M , Hikosaka O : Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 2007, 447:1111–1115. [PubMed: 17522629]
58. Ji H , Shepard PD : Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABA(A) receptor-mediated mechanism. J Neurosci 2007, 27:6923–6930. [PubMed: 17596440]
59. Christoph GR , Leonzio RJ , Wilcox KS : Stimulation of the lateral habenula inhibits dopamine-containing neurons in the substantia nigra and ventral tegmental area of the rat. J Neurosci 1986, 6:613–619. [PubMed: 3958786]
60. Stamatakis AM , Stuber GD : Activation of lateral habenula inputs to the ventral midbrain promotes behavioral avoidance. Nat Neurosci 2012, 15:1105–1107. [PubMed: 22729176]
61. Lammel S , Hetzel A , Hackel O , Jones I , Liss B , Roeper J : Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. Neuron 2008, 57:760–773. [PubMed: 18341995]
62. Grace AA , Bunney BS : Intracellular and extracellular electrophysiology of nigral dopaminergic neurons — 1. Identification and characterization. Neuroscience 1983, 10:301–315. [PubMed: 6633863]
63. Bunney BS , Aghajanian GK , Roth RH : Comparison of effects of L-dopa, amphetamine and apomorphine on firing rate of rat dopaminergic neurones. Nat New Biol 1973, 245:123–125. [PubMed: 4518113]
64. Schultz W , Romo R : Responses of nigrostriatal dopamine neurons to high-intensity somatosensory stimulation in the anesthetized monkey. J Neurophysiol 1987, 57:201–217. [PubMed: 3559672]
65. Aebischer P , Schultz W : The activity of pars compacta neurons of the monkey substantia nigra is depressed by apomorphine. Neurosci Lett 1984, 50:25–29. [PubMed: 6493628]
66. Mirenowicz J , Schultz W : Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. Nature 1996, 379:449–451. [PubMed: 8559249]
67. Brischoux F , Chakraborty S , Brierley DI , Ungless MA : Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. Proc Natl Acad Sci 2009, 106:4894–4899. [PubMed: 19261850]
68. Lerner TN , Shilyansky C , Davidson TJ , Evans KE , Beier KT , Zalocusky KA , Crow AK , Malenka RC , Luo L , Tomer R et al.: Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. Cell 2015, 162:635–647. [PubMed: 26232229]
69. Fiorillo CD : Two dimensions of value: dopamine neurons represent reward but not aversiveness. Science 2013,341:546–549. [PubMed: 23908236]
70. Fiorillo CD , Yun SR , Song MR : Diversity and homogeneity in responses of midbrain dopamine neurons. J Neurosci 2013, 33:4693–4709. [PubMed: 23486943]
71. Fiorillo CD , Song MR , Yun SR : Multiphasic temporal dynamics in responses of midbrain dopamine neurons to appetitive and aversive stimuli. J Neurosci 2013, 33:4710–4725. [PubMed: 23486944]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

72. Matsumoto M , Hikosaka O : Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 2009, 459:837–841. [PubMed: 19448610]
73. Menegas W , Babayan BM , Uchida N , Watabe-Uchida M : Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. Elife 2017:6.
74. Howe MW , Dombeck DA : Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature 2016, 535:505–510. [PubMed: 27398617]
75. Schultz W , Ruffieux A , Aebischer P : The activity of pars compacta neurons of the monkey substantia nigra in relation to motor activation. Exp Brain Res 1983, 51:377–387.
76. Schultz W : Dopamine reward prediction-error signalling: a two-component response. Nat Rev Neurosci 2016, 17:183–195. [PubMed: 26865020]
77. Nomoto K , Schultz W , Watanabe T , Sakagami M : Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. J Neurosci 2010, 30:10692–10702. [PubMed: 20702700]
78. Chen TW , Wardill TJ , Sun Y , Pulver SR , Renninger SL , Baohan A , Schreiter ER , Kerr RA , Orger MB , Jayaraman V et al.: Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 2013, 499:295–300. [PubMed: 23868258]
79. Kobayashi S , Schultz W : Reward contexts extend dopamine signals to unrewarded stimuli. Curr Biol 2014, 24:56–62. [PubMed: 24332545]
80. Sheafor PJ , Gormezano I : Conditioning the rabbit's (*Oryctolagus cuniculus*) jaw-movement response: US magnitude effects on URs, CRs, and pseudo-CRs. J Comp Physiol Psychol 1972, 81:449–456. [PubMed: 4649185]
81. Morris G , Arkadir D , Nevet A , Vaadia E , Bergman H : Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. Neuron 2004, 43:133–143. [PubMed: 15233923]
82. Shen W , Flajolet M , Greengard P , Surmeier DJ : Dichotomous dopaminergic control of striatal synaptic plasticity. Science 2008, 321:848–851. [PubMed: 18687967]
83. Tai LH , Lee AM , Benavidez N , Bonci A , Wilbrecht L : Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. Nat Neurosci 2012, 15:1281–1289. [PubMed: 22902719]
84. Brzosko Z , Zannone S , Schultz W , Clopath C , Paulsen O : Sequential neuromodulation of Hebbian plasticity offers mechanism for effective reward-based navigation. Elife 2017:6.
85. Marr D : Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Henry Holt and Company; 1983.
86. Krakauer JW , Ghazanfar AA , Gomez-Marin A , MacIver MA , Poeppel D : Neuroscience needs behavior: correcting a reductionist bias. Neuron 2017, 93:480–490. [PubMed: 28182904]
87. Tobler PN , Fiorillo CD , Schultz W : Adaptive coding of reward value by dopamine neurons. Science 2005, 307:1642–1645. [PubMed: 15761155]
88. Fiorillo CD , Tobler PN , Schultz W : Discrete coding of reward probability and uncertainty by dopamine neurons. Science 2003, 299:1898–1902. [PubMed: 12649484]
89. Nakahara H , Itoh H , Kawagoe R , Takikawa Y , Hikosaka O : Dopamine neurons can represent context-dependent prediction error. Neuron 2004, 41:269–280. [PubMed: 14741107]
90. Kobayashi S , Schultz W : Influence of reward delays on responses of dopamine neurons. J Neurosci 2008, 28:7837–7846. [PubMed: 18667616]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

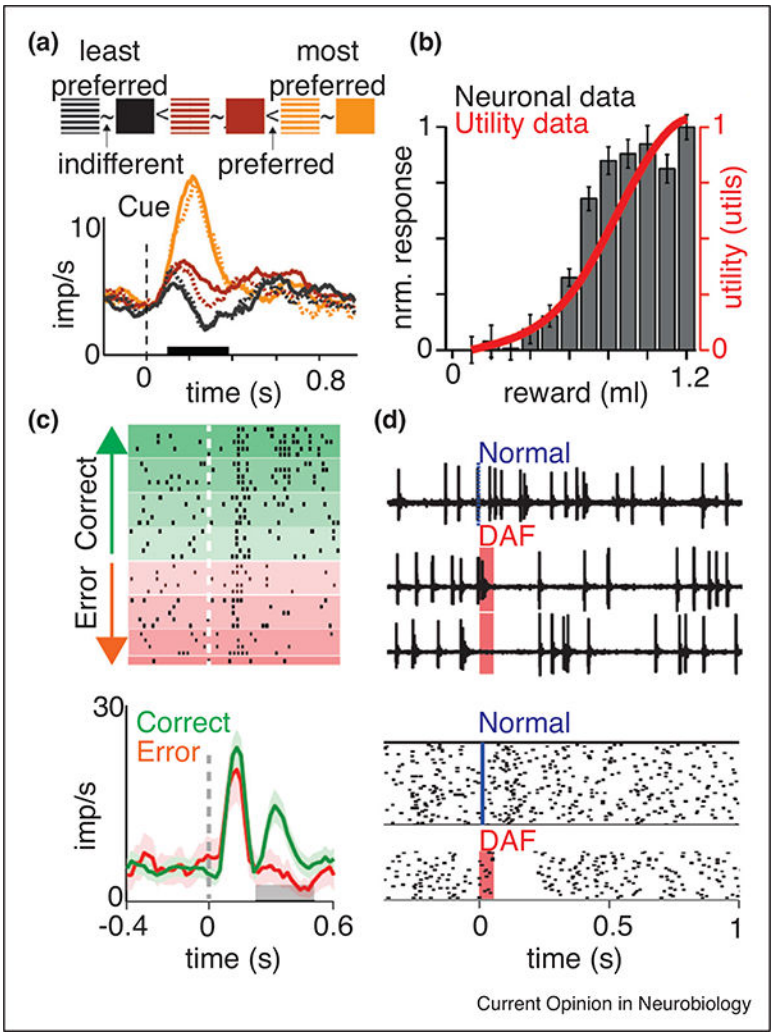

Figure 1:

Dopamine neurons reflect the behavioral computations of value. **(a)** Dopamine neurons represent a common scale of value. Monkeys indicated preference between different reward types by making choices. Orange and brown boxes represent the CSs that were associated with the different rewards, and that the monkeys chose between. ‘Greater than’ symbols indicate more preferred, whereas tildes indicate choice indifference. Peri-Stimulus Time Histogram (PSTH) of dopamine responses to onset of visual reward-predicting CS. Individual PSTH are color and dash coded according to the CS. Dopamine responses were largest for the most preferred rewards, smallest for the least preferred rewards, and indistinguishable for rewards that the monkey was indifferent between. This figure was modified and reproduced with permission from Ref. [20]. **(b)** The mathematical relationship between subjective value (utility) and physical reward size was described by an S-shaped function (red line). Grey bars indicate dopamine responses to unpredicted rewards that varied in magnitude between 0.1 ml and 1.2 ml in 0.1 ml increments. Error bars are SEM across 16 neurons. This figure was modified and reproduced with permission from Ref. [23]. **(c)** Raster plot (top) and PSTH (bottom) of one dopamine neuron in response to the onset of a RDM stimulus. Data are divided according to the accuracy of the subsequent choice.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Dopamine neurons were more active on trials when the monkey chose correctly (green), rather than incorrectly (red). Numbers along the side of the raster plot indicate RDM coherence. Shaded error bars on PSTH are SEM across trials. This figure was modified and reproduced with permission from Ref. [19]. **(d)** Dopamine neurons are silenced by distorted audio feedback (DAF) during bird song learning. Voltage traces (top) and raster plots (bottom) around normal ('Normal') and distorted ('DAF') audio feedback. This figure was modified and reproduced with permission from Ref. [18••].