HHS Public Access

Author manuscript

Annu Rev Neurosci. Author manuscript; available in PMC 2019 September 03.

Published in final edited form as:

Annu Rev Neurosci. 2017 July 25; 40: 373–394. doi:10.1146/annurev-neuro-072116-031109.

Neural circuitry of reward prediction error

Mitsuko Watabe-Uchida[1,*], **Neir Eshel**[1,2,*], **Naoshige Uchida**[1]

[1]Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, MA 02138

[2]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305

Abstract

Dopamine neurons facilitate learning by calculating reward prediction error, or the difference between expected and actual reward. Despite two decades of research, it remains unclear how dopamine neurons make this calculation. Here we review studies that tackle this problem from a diverse set of approaches, from anatomy to electrophysiology to computational modeling and behavior. Several patterns emerge from this synthesis: that dopamine neurons themselves calculate reward prediction error, rather than inherit it passively from upstream regions; that they combine multiple separate and redundant inputs, which are themselves interconnected in a dense recurrent network; and that despite the complexity of inputs, the output from dopamine neurons is remarkably homogeneous and robust. The more we study this simple arithmetic computation, the knottier it appears to be, suggesting a daunting (but stimulating) path ahead for neuroscience more generally.

Introduction

The brain is a prediction-making machine. Is that bobbing circle in the distance a human face? Whose face is it? How quickly will she arrive? For every piece of sensory information, our brains use stored patterns to generate a series of predictions. For each of these predictions, an outcome is ultimately experienced. The difference between prediction and outcome is the prediction error, which is thought to be a fundamental way that the brain learns from experience. If the error is small, there is no need to learn. If the error is large, however, the prediction must be updated. In this way, the brain ensures more optimal predictions in the future.

Predictive coding—the idea that the brain generates hypotheses, which are then tested against sensory evidence—has been discussed in a multitude of contexts, from visual processing to motor learning, cerebellum to cortex, simple organisms like electric fish to complex human diseases like schizophrenia (Friston, 2012; Rao and Ballard, 1999). Here we discuss one type of prediction error—reward prediction error—and one circuit that encodes it: the dopamine circuit. Synthesizing classic and recent findings, we develop a model for

mitsuko@mcb.harvard.edu.

*Equal contribution

how dopamine neurons calculate reward prediction error and how this signal is broadcast to the rest of the brain.

Reward prediction error coding by dopamine neurons

Dopamine and reward prediction error: an introduction

The idea that reward prediction errors help guide learning emerged first in psychology, with seminal work by Bush, Mosteller, Kamin, Rescorla, and Wagner, among others (Bush and Mosteller, 1951; Kamin, 1969; Rescorla and Wagner, 1972). Exploiting intricate behavioral tasks, these pioneers discovered that simple repetition was not always enough for animals to form a durable association between stimuli and reward. To ensure learning, surprise (i.e., an outcome different than expected) was critical. Later, this concept was appropriated by computer scientists, who created prediction-error algorithms to optimize how a computerized agent explores an unknown environment (Sutton and Barto, 1998). Indeed, prediction errors continue to play a role in even the most advanced computational algorithms, such as the one that recently mastered the game of Go (Silver et al., 2016).

In the mid-1990s, these models of learning gained a neurobiological flare when Wolfram Schultz and colleagues demonstrated the remarkable similarity between dopamine neuron firing rates and the reward prediction error signal (Mirenowicz and Schultz, 1994; Montague et al., 1996; Schultz et al., 1997; Waelti et al., 2001). When monkeys receive unexpected reward, dopamine neurons fire a burst of action potentials. If the monkeys learn to expect reward, that same reward no longer triggers a dopamine response. Finally, if an expected reward is omitted, dopamine neurons pause their firing at the exact moment reward is expected (Hollerman et al., 1998). Together, these results suggest that dopamine neurons signal the difference between the reward an animal expects to receive and the reward it actually receives. When reward is greater than expected, dopamine neurons fire; when reward is the same as expected, there is no response; when reward is less than expected, activity is suppressed.

Over the past 20 years, numerous electrophysiological and electrochemical recordings have confirmed and elaborated these results, investigating the properties of dopamine prediction errors and how these signals might facilitate learning in the brain (Glimcher, 2011; Schultz, 2013, 2016a). Activities consistent with reward prediction errors have been demonstrated in monkeys (Bayer and Glimcher, 2005; Enomoto et al., 2011), rats (Day et al., 2007; Flagel et al., 2011; Hart et al., 2014; Oleson et al., 2012; Roesch et al., 2007; Stuber et al., 2008) and humans (D'Ardenne et al., 2008), and appear to faithfully encode various features that determine reward value: probability (Fiorillo et al., 2003), magnitude (Bayer and Glimcher, 2005; Bayer et al., 2007; Tobler et al., 2005), timing (Fiorillo et al., 2008; Hollerman et al., 1998; Kobayashi and Schultz, 2008), and even subjective preference (Lak et al., 2014).

One limitation in the above electrophysiological studies is that the identification of dopamine neurons was based on indirect physiological properties such as wide spike waveforms (Grace and Bunney, 1983; Schultz, 1986; Ungless and Grace, 2012). These criteria are not always reliable (Lammel et al., 2008; Margolis et al., 2006). To circumvent these problems, Cohen et al. (2012) used optogenetics (Boyden et al., 2005; Lima et al.,

2009) to definitively identify dopaminergic neurons while recording in the ventral tegmental area (VTA). The authors tagged dopamine neurons with the light-sensitive cation channel, channelrhodopsin-2 (ChR2). Then, at the beginning and end of each recording session, they delivered pulses of blue light through a fiber optic, directly into the region of VTA being recorded. Dopamine neurons identified using this approach showed reward-prediction error-related activities, confirming previous studies (Cohen et al., 2012; Tian and Uchida, 2015) (Figure 1A, B). More recently, Stauffer et al. (2016) applied this technique in a non-human primate, opening the possibility to perform rigorous identifications in non-human primates.

Arithmetic of dopamine prediction errors

How do dopamine neurons calculate reward prediction error? Reinforcement learning models long assumed that dopamine neurons perform subtraction, i.e., reward prediction error = actual reward – expected reward. However, this arithmetic was never explicitly tested against alternative possibilities.

To explore the nature of the computation, it is helpful to think of dopamine neurons as if they were sensory neurons—but instead of encoding the decibels of a sound, for example, they encode the extent of prediction error. Sensory neurons transform external information into an internal variable: firing rate. If a neuron is tuned to a particular stimulus (e.g., sound), then the more intense the stimulus, the more that neuron will fire. Often a neuron responds minimally to low intensities, increases its response with a certain slope, and then saturates at an asymptotic level, resembling a sigmoid. This input-output function, however, is not fixed: the threshold, slope and saturation level are all modulated by the nature of sensory inputs and the context in which they are presented. These changes in response functions are examples of “neuronal arithmetic” and are thought to be essential for the brain to process behaviorally-relevant sensory information (Silver, 2010; Uchida et al., 2013). In particular, the fields of vision (Atallah et al., 2012; Lee et al., 2012; Williford and Maunsell, 2006; Wilson et al., 2012) and olfaction (Kato et al., 2013; Miyamichi et al., 2013; Olsen et al., 2010; Papadopoulou et al., 2011) abound with examples of neuronal arithmetic, with a rich modeling (Ayaz and Chance, 2009; Holt and Koch, 1997; Murphy and Miller, 2003) and experimental (Cardin et al., 2008; Chance et al., 2002; Olsen et al., 2010; Shu et al., 2003) literature exploring the biophysical mechanisms that might underlie it. Until recently, this type of analysis was lacking in the dopamine field.

Although many models of reinforcement learning assumed subtraction, division is equally possible, and in fact much more commonly found in other systems in the brain (Silver, 2010). To find out which computation dopamine neurons use, Eshel et al. (2015) recorded from optogenetically-identified dopamine neurons in the lateral VTA as mice performed a simple classical conditioning task. Mice received various sizes of water reward: sometimes these rewards were delivered unexpectedly, in the absence of any cue, and sometimes they were preceded by an odor cue. By using a number of different reward sizes, the authors inferred the full dose-response function for dopamine neurons, i.e., the number of spikes that dopamine neurons fired for various rewards (Figure 2A). They then compared this response function when reward was expected (because of a cue) or unexpected (Figure 2B). The authors found that expectation reduces the phasic reward responses of dopamine neurons in

a purely subtractive fashion (Figure 2C). No matter the size of reward, a given level of expectation caused a consistent decrease in dopamine responses. This is an unusual computation in the brain but one that is consistent with classic reinforcement learning models.

These analyses were done by averaging over all recorded dopamine neurons. As a population, then, dopamine neurons use simple subtraction. How do individual neurons make this computation? In a subsequent paper, Eshel et al. (2016) determined the full prediction error functions for each individual dopamine neuron in the lateral VTA, and assessed how these functions related to each other. They found remarkable homogeneity among neurons. Each dopamine neuron appeared to use the same function, just scaled up or down (Figure 2D). Thus, dopamine neurons provide an ideal broadcast signal: similar enough from neuron to neuron that downstream targets could decode the same information regardless of the subset of dopamine neurons that they contact. Such robust encoding had long been inferred (Fiorillo et al., 2013; Glimcher, 2011; Schultz, 2013), but the quantitative relationship between individual neurons had never been demonstrated.

Functions of dopamine prediction error signals

Dopamine has long been thought to be a key regulator of reward-based learning (Wise and Rompre, 1989). The above findings indicate that dopamine can drive learning through signaling prediction error. Recent experimental findings using newer techniques have reinforced this idea, although they by no means exclude other functions of dopamine (Wise, 2004). With optogenetics, it became possible to manipulate dopamine neurons with the temporal and genetic precision required to probe their causal effect on learning. Using this technique, it has been shown that activation or inhibition of dopamine neurons is sufficient to reinforce a behavior positively (Tsai et al., 2009; Witten et al., 2011) or negatively (Danjo et al., 2014; Tan et al., 2012; van Zessen et al., 2012), respectively. Importantly, using the so-called ‘blocking’ paradigm, Steinberg et al. (Steinberg et al., 2013) demonstrated that phasic increases in dopamine firing were sufficient for prediction-error-induced learning. Stimulating dopamine neurons at the time of an expected reward caused rats to learn an association with an otherwise ‘blocked’ cue, presumably through positive prediction errors. Conversely, Chang et al. (Chang et al., 2016) used the paradigm of Pavlovian over-expectation to show that phasic decreases in dopamine responses were also sufficient to produce learning, this time through negative prediction errors. Together, these results demonstrate that dopamine prediction errors regulate learning in both positive and negative directions.

Complexities in the dopamine signal

So far we have seen evidence favoring a very simple story: dopamine neurons calculate prediction error by subtracting expected from actual reward, and then broadcast this signal accurately and consistently to target regions in the brain, promoting learning from trial and error. This simple story, however, belies very important complexities in the nature of this signal.

The findings above on the homogeneity of dopamine prediction error functions fit with a classic literature showing that dopamine neurons have stereotyped electrophysiological properties (Grace and Bunney, 1983), electrically couple with each other (Vandecasteele et al., 2005), and coordinate their *in vivo* firing rates (Joshua et al., 2009; Kim et al., 2012; Morris et al., 2004; Schultz, 1998). However, it is now clear that dopamine neurons are not all the same (Bromberg-Martin et al., 2010; Volman et al., 2013). A host of recent studies have shown diversity in every aspect of dopamine neurons: from their physiology (Margolis et al., 2006; Neuhoff et al., 2002), to their connectivity (Lammel et al., 2008, 2012; Watabe-Uchida et al., 2012), and even their genetic profiles (Blaess et al., 2011; Haber et al., 1995).

Importantly for this review, some dopamine neurons do not faithfully calculate prediction error in the first place. Instead, they increase their firing to both rewarding and aversive events (Fiorillo et al., 2013; Horvitz, 2000; Joshua et al., 2008; Lerner et al., 2015; Matsumoto and Hikosaka, 2009a). Rather than encoding the difference between actual and predicted outcome, these neurons might encode “motivational salience” or the absolute value of this difference (Matsumoto and Hikosaka, 2009a). Mostly found in more lateral regions of the midbrain, particularly the substantia nigra pars compacta (SNc) (presumably projecting to the dorsal striatum), these neurons may be important in marking behaviorally important stimuli rather than in updating value assignments (Matsumoto and Hikosaka, 2009a) (but see Fiorillo et al., 2013).

While the nature of these ‘non-canonical’ dopamine signals remains to be further characterized, a preponderance of evidence suggests that dopamine in the nucleus accumbens (Acb) encodes prediction error signals relatively faithfully in simple tasks (Flagel et al., 2011; Hart et al., 2014; Stuber et al., 2008; Wenzel et al., 2015). Roitman et al. (2008) reasoned that release of dopamine in Acb in response to aversive stimuli may be due to confounding factors such as the difference in sensory modality or intensity. To control for these differences, they used sucrose and quinine solutions for appetitive and aversive stimuli, respectively. They found that these stimuli caused opposite responses: sucrose increased and quinine decreased dopamine release in the Acb, suggesting that at least the majority of dopamine neurons projecting to the Acb are inhibited by aversive stimuli (Roitman et al., 2008). Another study showed that Acb dopamine could be elevated when an animal successfully avoided an aversive event, suggesting that some of the excitation to aversive stimuli could be regarded as a “safety” signal (Oleson et al., 2012; Wenzel et al., 2015).

Another important factor to be considered is that dopamine neurons’ responses may depend on reward context (Kobayashi and Schultz, 2014). Previous studies that recorded from optogenetically-identified dopamine neurons typically found bi-phasic responses (short-latency, transient excitation followed by inhibition) to airpuff or airpuff-predictive cues (Cohen et al., 2012; Tian and Uchida, 2015) (Figure 1A, B). A recent study, however, found that most dopamine neurons in the lateral VTA show pure inhibition to cues predicting aversive airpuffs in a certain task condition (Matsumoto et al., 2016). The difference is due to reward context: the short-latency, transient excitation appears in high reward contexts but disappears in low reward contexts (Figure 1C). Therefore, some of the excitatory responses to aversive events can be due to the effect of high-reward contexts, which are commonly used in recording experiments (Cohen et al., 2012; Tian and Uchida, 2015). Schultz recently

posited that there are two components of the phasic dopamine signal (Schultz, 2016b). An initial stage (the first ~200 ms) is unselective, detecting physical salience, rather than prediction error. Matsumoto et al. (2016), discussed above, showed that this initial response is more vulnerable to context-dependent modulations. Later on, from 200–400ms after stimulus onset, dopamine neurons show a more fine-grained prediction-error response (Schultz, 2016b), which could be obscured by the initial response in certain conditions such as in high reward contexts (Fiorillo, 2013; Matsumoto et al., 2016).

Besides prediction errors, some dopamine neurons also appear to encode movement-related information (Howe and Dombeck, 2016; Jin and Costa, 2010; Kim et al., 2015; Parker et al., 2016). Howe and Dombeck (2016), for example, found that dopamine axons projecting to the dorsal striatum transiently elevate their activity around the onset of locomotor movements (Also see, Jin and Costa, 2010). Other studies found that dopamine activity can be modulated by direction of movement (Kim et al., 2015; Parker et al., 2016). Furthermore, more sustained or ramping dopamine signals have been found when the animal is engaged in certain task conditions (Hamid et al., 2016; Howe et al., 2013; Takahashi et al., 2011) (but see Gershman, 2014).

The physiology of dopamine signal appears more complex the more it is studied. As discussed above, there is certainly diversity of responses, pointing to the importance of characterizing dopamine responses with the firm knowledge of their neurochemical identity as well as their projection targets. At the same time, recent experiments have provided stronger evidence that RPE constitutes a core component of the dopamine signal. To understand the circuit mechanism, it is, therefore, helpful to return to the simplest version of RPE, and design experiments to understand how this simple arithmetic can be instantiated by a neural network. What are the inputs and how are they combined?

Computation of dopamine prediction errors

Models of the prediction error computation

As discussed above, the activity of dopamine neurons can be approximated by simple equations (Eshel et al., 2015; Schultz et al., 1997). Multiple theories have proposed how RPE could be computed in the brain (Brown et al., 1999; Hazy et al., 2010; Houk and Davis, 1995; Joel et al., 2002; Kawato and Samejima, 2007; Schultz, 1998; Schultz et al., 1997; Stuber et al., 2008; Tan and Bullock, 2008; Vitay and Hamker, 2014). Although there are numerous differences between these models, most boil down to three components: regions that encode expectation, region that encode actual reward, and the subtraction of these inputs at a common downstream target, usually dopamine neurons themselves. This system nicely explains how dopamine neurons respond to unexpected or expected reward. However, important mysteries remain, including how dopamine neurons become excited by reward-predicting cues or inhibited by omission of expected reward. How is reward information transferred from the reward itself to earlier stimuli? And how does the system learn the precise timing of reward?

To answer these questions, the temporal difference (TD) learning model posits that the cue and omission responses both emerge from the same inputs (Figure 3A) (Houk and Davis,

1995; Kawato and Samejima, 2007; Morita et al., 2013; Schultz et al., 1997; Sutton and Barto, 1998). In the simplest version, there are two sustained expectation signals, both of which rise at the cue predicting reward and fall at the time of reward itself. One of these signals [V(t)] is excitatory while the other [V(t+1)] is inhibitory. Importantly, the inhibitory signal is slightly temporally shifted, so that it begins and ends later than the excitatory signal. By summing these signals, dopamine neurons would show phasic excitation at reward predicting cues and inhibition at reward omission (Figure 3A). In other words, RPE is generated by taking the derivative of reward prediction.

TD models are successful in explaining many aspects of dopamine RPEs and provide a link between the dopamine reward circuit and machine learning mechanisms. However, this does not prove that RPEs are calculated exactly as the equation predicts. It remains unclear whether the precise time-shift between V(t) and V(t+1) is realistic or what neural substrates could underlie such signals. Thus, another group of models posits that there are multiple inputs that separately represent reward prediction at the CS and at the US (Figure 3B, C) (Brown et al., 1999; Contreras-Vidal and Schultz, 1999; Hazy et al., 2010; O'Reilly et al., 2007; Tan and Bullock, 2008; Vitay and Hamker, 2014). One set of inputs excites dopamine neurons at the reward-predicting cue and another inhibits reward responses when reward is predicted. The latter system may also produce the dopamine dip when reward is omitted, or this could be accomplished by a third input. Using separate systems is advantageous in the sense that it provides flexibility to build complex features, such as independent learning rates or variable timing.

Finally, several authors have suggested that prediction error is not calculated by dopamine neurons at all, but rather in upstream areas such as the rostromedial tegmental nucleus (RMTg) (Jhou et al., 2009) or lateral habenula (lHb) (Hong et al., 2011; Matsumoto and Hikosaka, 2007, 2009b). The information is then relayed to dopamine neurons. In this view, the dopamine neurons are passive conveyers of information, rather than active comparators of actual and predicted reward.

Anatomy of dopamine inputs

To distinguish between these models and understand how dopamine RPE signals are generated, we start with anatomy. Which brain areas actually project to dopamine neurons? Using retrograde tracers, Zahm and colleagues (Geisler and Zahm, 2005; Geisler et al., 2007) systematically examined projections to VTA and found extensive sources of inputs, particularly around the medial forebrain bundle. Importantly, the authors proposed that brain areas that project directly to VTA tend to also project indirectly to VTA. For example, there is a loop from PFC to nucleus accumbens (Acb) to ventral pallidum (VP) to lateral hypothalamus (LH) to VTA, suggesting an interconnected anatomical network for the regulation of dopamine neurons.

Although informative, conventional retrograde tracers cannot distinguish cell types of the target areas. Since the VTA consists of dopamine neurons, GABA neurons, and glutamate neurons, some of the areas that project to VTA may not in fact project to dopamine neurons. To overcome this barrier and label monosynaptic inputs to dopamine neurons specifically, Watabe-Uchida et al. (Watabe-Uchida et al., 2012) applied a modified rabies virus system

(Wickersham et al., 2007). Using this system, the authors comprehensively mapped inputs to dopamine neurons and found many brain areas that project directly to dopamine neurons (Figure 4A, B). Comparing inputs to dopamine neurons in VTA and SNc, the authors proposed that lateral orbitofrontal cortex (OFC) and LH are the major excitatory inputs to VTA, while sensorimotor cortex and subthalamic nucleus are the major excitatory inputs to SNc. On the other hand, the ventral and dorsal nuclei in the basal ganglia (the Acb and VP, versus the dorsal striatum, globus pallidus, entopeduncular nucleus and substantia nigra reticulata) provide the major inhibitory inputs to dopamine neurons in VTA and SNc, respectively.

The above study mapped inputs to dopamine neurons in VTA and SNc regardless of where those dopamine neurons project. However, it is well known that dopamine neurons within a given region may project to a diverse array of targets. Thus, the previous study might have observed so many monosynaptic inputs because the dopamine neurons were themselves diverse. Recent studies tackled this problem by mapping monosynaptic inputs to subpopulations of dopamine neurons that project to specific brain areas (Beier et al., 2015; Lerner et al., 2015; Menegas et al., 2015). These studies found that even when the projection targets of dopamine neurons were specified, there were still monosynaptic inputs from a large number of brain areas. Menegas et al (2015) found that, for 7 of the 8 examined subpopulations of dopamine neurons, monosynaptic inputs were largely overlapping. Unexpectedly, however, dopamine neurons that projected to the “tail” of the striatum (TS, the most posterior part of the striatum) received a different set of inputs, suggesting that the function of TS-projecting dopamine neurons might be different from most dopamine neurons. Assuming that most dopamine neurons encode RPE, the inputs specific to TS-projecting dopamine neurons can be excluded from the list of inputs needed for RPE. The brain areas that remain—those that appear to provide major inputs to RPE-encoding dopamine neurons—are the LH, the ventral and dorsal striatum, the lateral preoptic area, and the VP. With this structural information in hand, the next question is: what information does each area send to dopamine neurons?

Electrophysiology of inputs

Decades of recordings have provided important hints on candidate brain areas for RPE computations. For example, the lateral hypothalamus (LH) is known to encode reward information such as taste (Ono et al., 1986). Combined with the fact that responses to reward are modulated by internal states such as hunger (Burton et al., 1976), these results suggest that the LH may encode subjective values, which could be sent to dopamine neurons directly or indirectly. On the other hand, striatal neurons respond to reward-predicting cues, often showing sustained excitation (Hikosaka and Sakamoto, 1986; Schultz et al., 1993). Together with the fact that the overwhelming majority of striatal neurons are inhibitory, this response pattern makes the striatum a good candidate for providing the expectation signal. Because the striatum is the main projection target of dopamine neurons, reciprocal connections between the striatum and dopamine neurons would make learning straightforward: the striatum sends predicted value and dopamine neurons return prediction error. Further, direct and multi-synaptic pathways from the striatum to dopamine neurons imply several potential mechanisms to produce RPE.

Although electrophysiology can be incredibly informative, there are important pitfalls in the interpretation of these results. First, although recording experiments can find interesting activity in a given brain area, other areas may have the same responses. Hence the need for the systematic study of many brain regions, ideally simultaneously. Second, neurons that seem to encode information relevant to the task at hand are often intermingled with neurons that show other types of activity. We seldom know which information is sent to a specific downstream brain target. Third, even if we know the projection target of the neurons, the target brain areas are themselves diverse, and it can be unclear to which specific type of neuron the information is going. In the case of reward prediction errors, the relevant question is which brain areas send information about actual and expected reward to dopamine neurons.

To directly answer this question, Tian et al. (2016) established an awake recording system that combined optogenetics with the modified rabies virus. While mice performed simple classical conditioning tasks, the authors recorded extracellular activity of monosynaptic inputs to dopamine neurons in 7 input areas: dorsal striatum (DS), nucleus accumbens (Acb), VP, LH, subthalamic nucleus, RMTg, and pedunculopontine tegmental nucleus (PPTg) (Figure 4C). Surprisingly, there were input neurons in all 7 recorded areas that encoded either actual reward or expectation. In fact, many single input neurons were modulated by both actual reward and expectation. Thus, information relevant to reward prediction error had already been combined—at least in part—in input neurons. Importantly, however, very few input neurons showed complete reward-prediction-error signals. Thus, dopamine neurons receive a spectrum of information, including pure reward, pure expectation, mixed reward and expectation, partial RPE, and in rare cases, complete RPE, all from multiple brain areas. This then gets funneled into a pure RPE signal in dopamine neurons. In other words, the brain seems to perform the RPE computation redundantly, in multiple layers.

At first glance, these results are puzzling. The brain appears to solve the problem in a very inefficient way. However, a simple model helps explain the findings. Tian et al. (2016) first created a linear model to reconstruct dopamine activity using the activity of input neurons from all 7 areas. They found that a weighted sum of inputs could easily reconstruct dopamine activity. Indeed, even if an entire brain area were removed from the analysis, the remaining inputs could still reconstruct dopamine RPEs. The same is true even if the weights for each input were totally shuffled: the resulting output sill captured aspects of RPE signals. On the other hand, if recordings from other neurons in these regions, which were not identified as inputs, were used for the model, the reconstructions became less accurate. These results suggest that the identity of the inputs is important, even if the precise weights between inputs and dopamine neurons are not. Thus, far from inefficient, the presence of mixed information appears to be a convenient, robust, and ready-to-use format for dopamine neurons to compute RPE.

One open question, then, is whether inputs to dopamine neurons are redundant or specialized. The data presented above suggests the former. Another open question is the importance of excitatory versus inhibitory projections to dopamine neurons. For this, Tian et al. (2016) identified input neurons that could discriminate conditioned stimuli based on

probability of reward, and whose responses were fast enough to account for the dopamine CS response. They found that all input neurons that met these criteria were excited by reward cues. Because dopamine neurons are also excited by reward cues, these results suggest that excitatory inputs (particularly in VP, LH, and PPTg) likely cause dopamine phasic responses to CS. In other words, disinhibition (i.e. inhibition of inhibitory inputs) appears to play a very limited role. In fact, this pattern held true for responses to aversive stimuli as well: most monosynaptic inputs to dopamine neurons were excited by aversive stimuli, implying that the suppression in dopamine neurons must be due to direct inhibition (e.g., from inhibitory neurons in Acb or RMTg), rather than reduced excitation.

The presence of both direct excitation and direct inhibition implies that a combination of inputs must determine the dopamine RPE response, rather than variations in a single type of input. Further evidence for this claim comes from analyzing both the time-course and amplitude of dopamine responses. Matsumoto et al. (2016) showed that dopamine excitation to reward-predicting cues occurs faster than inhibition to aversion-predicting cues, implying different inputs for each process. In addition, as mentioned above, Eshel et al. (2016) found that response functions to reward were remarkably uniform from neuron to neuron. It turns out that response functions to aversive stimuli are similarly uniform (Matsumoto et al., 2016). However, responses to reward are not correlated with responses to aversive stimuli (Matsumoto et al., 2016). Furthermore, habenula lesions showed that the dopamine dip during reward omission disproportionately depends on the function of the lateral habenula (lHb), while the dip to aversive stimuli does not (Tian and Uchida, 2015). These findings suggest the presence of multiple separate inputs determining excitation and inhibition in dopamine neurons, raising the possibility that RPE computations consist of multiple mechanisms.

Local connections in VTA

Much of the anatomic and physiologic work discussed so far has focused on long-range projections to dopamine neurons. However, dopamine neurons in VTA are surrounded locally by GABA neurons and glutamate neurons, both of which send projections to their dopaminergic neighbors (Sesack and Grace, 2010). Because of the vicinity, local connections may have particularly strong effects on dopamine activity.

To dissect the different roles of neurons in VTA, Cohen et al. (2012) recorded from VTA while mice performed classical conditioning tasks. They found three types of activity: type 1 resembled reward prediction error, with phasic activity to cues and rewards; type 2 resembled reward expectation, with a ramping cue response proportional to expected reward; and type 3 was a mirror image of type 2, with downward-sloping activity dependent on the magnitude of expected reward. Using the optogenetic identification method described above, the authors showed that while identified dopamine neurons were all type 1 (reward prediction error), identified GABA neurons were type 2, signaling reward expectation (Figure 1A).

At this moment, it is not clear whether type 3 neurons are GABAergic or glutamatergic. Of note, RMTg neurons are located together with dopamine neurons at the boundary between VTA and RMTg. Because most neurons in RMTg are GABA neurons and both type 3 and

RMTg neurons show inhibition in response to reward cues (Hong et al., 2011; Jhou et al., 2009), it is possible that they are actually one and the same. In this case, type 3 neurons would be GABA neurons. However, there are difference in activity between type 3 neurons and RMTg neurons. Whereas many VTA type 3 neurons show sustained inhibition to reward cues, and are not modulated by reward itself, most monosynaptic inputs from RMTg are modulated by both cues and rewards (Tian et al., 2016). Further classification—including experiments that tag glutamate neurons with ChR2—will be necessary to clarify the activity patterns of each cell type in VTA.

Functional studies: causality of inputs

Although many models assume that RPE is calculated in dopamine neurons, some have argued that dopamine neurons merely relay already-calculated RPEs. Neurons in the lateral habenula (lHb), for example, encode negative RPE (i.e., the mirror-image of dopamine RPE). It has therefore been proposed that they send RPE signals to dopamine neurons via GABAergic neurons in RMTg (Matsumoto and Hikosaka, 2007; Stephenson-Jones et al., 2016). If the lHb-RMTg system is the source of dopamine RPE, lesioning Hb should deplete RPE signals in dopamine neurons. Tian et al. (2015) tested this theory by lesioning the habenula while recording from optogenetically-identified dopamine neurons in the VTA. After lesions, dopamine neurons maintained their responses to reward and reward-predicting cues. Thus, consistent with the anatomy study above, the lHb-RMTg cannot be the only source of RPE. However, dopamine neurons did lose their inhibitory responses to reward omission after the habenula lesion, suggesting that there may be a specific function of these inputs. Further supporting the specificity of these inputs, dopamine neurons did not lose their responses to all aversive events; they maintained their responses to air puffs. Thus, lHb appears to be important for determining dopamine neurons' inhibition specifically to reward omission. In support of this idea, responses to reward omission were particularly vulnerable to changes in weights or input areas used in the linear combination of inputs discussed above (Tian et al., 2016). Perhaps the information important for reward omission arises in OFC (Feierstein et al., 2006) and then passes through the nucleus accumbens (Acb), entopeduncular nucleus (EP), lHb and RMTg before reaching dopamine neurons.

Beyond the dip to reward omission, one of the core features of RPE is that the response to reward is diminished when reward is expected. As discussed above, this reduction in reward response occurs through subtraction (Eshel et al., 2015). The obvious next question is: what inputs do dopamine neurons subtract? One possibility is VTA GABA neurons, which were shown to encode reward expectation (Figure 1A) (Cohen et al., 2012). Do dopamine neurons use this signal to calculate reward prediction error? Using optogenetics, Eshel et al. (2015) selectively excited or inhibited VTA GABA neurons while mice performed classical conditioning tasks, and determined the effect of this manipulation on putative dopamine responses. They found that stimulating VTA GABA neurons during the delay between the cue and the reward caused a subtraction of dopamine reward responses, mimicking the effect of reward expectation. Conversely, inhibiting VTA GABA neurons during this period increased dopamine responses to expected reward, as if reward were suddenly less expected. Finally, bilaterally stimulating VTA GABA neurons changed the animals' behavior, causing them to reduce their responses to laser-paired cues. This behavior is consistent with GABA

stimulation causing an over-exuberant prediction signal, which then led to reduced dopamine responses when reward actually came. Together, these results imply that VTA GABA neurons help put the 'prediction' in 'prediction error.' Of course, they may not be the only important inputs for reward expectation. It is important to know that VTA GABA neurons elevate their activity as soon as the reward-predictive cue is presented while dopamine neurons' baseline firing rates are not significantly inhibited (Figure 1A). This suggests that an intricate balance between GABA neuron activity and counteracting excitation, as proposed in TD models (Figure 3A), might underlie this process. Furthermore, Tian et al. (2016) demonstrated that other monosynaptic inputs have similar properties (Figure 4C). Although VTA GABA neurons' proximity and the density of their projections to dopamine neurons may give them an outsized role in RPE calculations, the RPE computation likely depends on inputs from diverse areas.

It remains to be determined where reward expectation is calculated in the first place or what drives the activity of VTA GABA neurons. Previous studies have shown that many neurons in the OFC and nucleus accumbens (Acb) change their activity depending on reward expectation. Takahashi and colleagues tested the role of these areas in producing RPE signals in dopamine neurons. First, lesions of lateral and ventral OFC in rats reduced putative dopamine neurons' ability to modulate their responses when the size or timing of reward was changed (Takahashi et al., 2011). Furthermore, Takahashi et al. (2016) found that after lesions of the Acb, putative dopamine neurons lost their ability to modulate their responses when the timing but not the size of reward was altered. These results suggest that OFC or Acb may play a role in the calculation of RPEs. However, these experiments depended on permanent lesions and used a learning paradigm to probe RPEs, making it difficult to dissociate whether altered RPE signaling was a direct effect of the lesion, or rather due to an impairment in the animal's ability to learn. Furthermore, it was not possible to examine whether the recorded neurons were indeed dopamine neurons, let alone how the activity of VTA GABA neurons was altered by the lesions. Studying the pathway by which OFC and Acb modulate the activity of dopamine neurons may provide important insights into how neural circuits compute RPEs.

Progress and future directions

The idea that dopamine neurons signal RPEs has revolutionized the study of reward processing and decision-making in the brain. In particular, the fact that dopamine responses can be well approximated by simple arithmetic equations has prompted many researchers to propose simple models that combine different variables in a single step. However, as more sophisticated anatomic and electrophysiologic data have become available, a complicated picture has emerged for the circuit underlying prediction errors. It is now clear, for example, that many brain areas project directly to VTA dopamine neurons (Watabe-Uchida et al., 2012). This is true not just for dopamine neurons as a whole but even for subsets of dopamine neurons defined by their projection targets (Beier et al., 2015; Lerner et al., 2015; Menegas et al., 2015). Moreover, there is striking interconnectivity between these input areas, more often resembling a recurrent neural network than a simple feedforward box-and-arrow diagram (Geisler and Zahm, 2005). From the standpoint of electrophysiology, presynaptic neurons or input neurons from all of these input areas show a diverse set of

responses (Tian et al., 2016). In even the simplest task, it is a rare input neuron that exhibits a pure response to either reward or expectation. Rather, information relevant for RPE computations is mixed and distributed throughout multiple regions. With such complexity, can we ever understand how the brain computes RPE?

Fortunately, despite the complexity, patterns have begun to emerge. The data presented above strongly support the centrality of dopamine neurons in the RPE calculation. Rather than receiving RPE signals passively from upstream regions such as lHb, dopamine neurons appear to be performing computations on their input (Eshel et al., 2015). Although most models assume pure inputs onto dopamine neurons, our work has demonstrated monosynaptic inputs that span from pure (such as reward information from VP, LH, and PPTg), to decidedly mixed (Tian et al., 2016). Ultimately, dopamine neurons are capable of combining these inputs in such a way to produce a remarkably homogeneous prediction error signal, ideal to broadcast to a wide variety of downstream targets (Eshel et al., 2016).

In terms of the mechanisms behind these computations, phasic increases in dopamine responses appear to be triggered by direct excitation, rather than disinhibition. Likewise, phasic decreases in dopamine responses appear to emerge from direct inhibition, rather than reduced excitation (although reduced excitation in PPTg may enhance prediction-dependent suppression of dopamine reward responses; see Tian et al. 2016).

Furthermore, unlike the simplest TD models, it does not appear that the CS and US responses are due to the same inputs. Instead, dopamine neurons appear to be combining multiple separate inputs, with different signals crucial for cue responses, reward responses, reward omission, and aversion. Indeed, even the dip in dopamine responses when reward is omitted appears to be due to a different input than the reduced but still positive reward response that dopamine neurons exhibit when reward is expected (Eshel et al., 2015; Tian and Uchida, 2015). And when it comes to aversive events, the inhibition in dopamine neurons appears to arise from a slower set of inputs than excitation to reward (Matsumoto et al., 2016).

In terms of brain regions involved in this circuit, there appears to be significant redundancy. Neurons in multiple regions all provide dopamine neurons with information relevant for RPEs. Interestingly, the relative weights of these inputs can change dramatically without affecting the final dopamine response (Tian et al., 2016). However, the inputs are not random —neurons that reside in input regions, but do not themselves project to dopamine neurons, cannot easily produce the dopamine RPE response (Tian et al., 2016). Based on a combination of anatomic and physiologic studies, key excitatory inputs include the lateral OFC, VP, lateral preoptic area and LH, while inhibitory inputs include VTA GABA neurons, the Acb, VP, and lateral preoptic area. Interestingly, VTA GABA neurons receive similar monosynaptic inputs as dopamine neurons (Beier et al., 2015). This sets up the possibility of feed-forward inhibition, allowing dopamine neurons to take derivatives, one of the major prediction of TD models.

Although studies have become increasingly specific, with analyses that target specific cell types and projection targets, most have focused on just one type of information flow: from

input areas to dopamine neurons. In the future, it will be crucial to analyze flow between input areas as well. For example, although excitatory inputs in multiple areas (VP, LH, and PPTg, to name a few) may trigger phasic responses to rewards and reward-predicting cues in dopamine neurons, VP may by the lynchpin, providing information not just to dopamine neurons but also to these other input regions. By hopping back through each node in the circuit, combining anatomy and physiology at every step, we may be fast approaching the solution to this mysterious and crucial circuit for computing RPEs.

The research discussed above has revealed the unprecedented complexity that underlies a seemingly simple arithmetic computation. This suggests some fundamental limitations that neuroscience faces more generally. First, the ultimate goal of neuroscience is to explain how neural circuits produce complex behavior. To this end, we often develop simple box-and-arrow models in which information flows from area A to B to C. It has also become a common practice to artificially activate a population of neurons and infer functions based on changes in behavior. Although these approaches can be useful, they often ignore far more complex connectivity between these areas and far more diverse activity within each area. The brain is a dynamical system that consists of many diverse and interconnected neurons. It is vital to consider whether these simplified views help our understanding or unintentionally hide essential features of neural circuits.

Second, many of the studies described above relied on monitoring the activity of neurons by recording action potentials. Monitoring spikes together with a neuron's cell type and connectivity has considerably enhanced our knowledge of neuronal function. Spikes are the main currency with which neurons communicate, and are therefore essential to understanding neuronal computations. However, spikes do not necessarily capture all the processes required for RPE computations. Importantly, RPE computations require comparing actual reward against what the animal expects. The latter requires memory about reward in a given context, and memory is likely be stored in synaptic weights or through intrinsic properties of neurons (Martin et al., 2000; Schultz et al., 1997). Spikes may not faithfully represent these memories because once a neuron fires a spike, the information propagates in neural circuits that affect other neurons' as well as its own activity. At present, our ability to monitor synaptic weights or intrinsic properties of neurons in behaving animals is quite limited. It is possible that the memory that supports RPE computations is stored in a pure fashion. Developing techniques that allow for monitoring synaptic weights, together with other variables (spikes, cell types, and connectivity), will advance our understanding.

Third, the results discussed above reinforce the importance of developing theories about how neural circuits with complex connectivity can perform simple computations robustly and accurately. How is the information stored, and how does it propagate? Are there computational advantages to redundant or distributed computations, versus simple box-and-arrow computations? As is the case in modern artificial neural networks (LeCun et al., 2015), it is often difficult to infer even simple operating principles by looking solely at the component parts. A crucial step is to develop theoretical frameworks that link global, top-down functions with how each component supports these functions. Only with such models can we hope to "understand" neural circuits, even one as seemingly basic as reward prediction errors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Reference

Atallah BV, Bruns W, Carandini M, and Scanziani M (2012). Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. Neuron 73, 159–170. [PubMed: 22243754]

Ayaz A, and Chance FS (2009). Gain Modulation of Neuronal Responses by Subtractive and Divisive Mechanisms of Inhibition. J. Neurophysiol 101, 958–968. [PubMed: 19073814]

Bayer HM, and Glimcher PW (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47, 129–141. [PubMed: 15996553]

Bayer HM, Lau B, and Glimcher PW (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. J. Neurophysiol 98, 1428–1439. [PubMed: 17615124]

Beier KT, Steinberg EE, DeLoach KE, Xie S, Miyamichi K, Schwarz L, Gao XJ, Kremer EJ, Malenka RC, and Luo L (2015). Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping. Cell 162, 622–634. [PubMed: 26232228]

Blaess S, Bodea GO, Kabanova A, Chanet S, Mugniery E, Derouiche A, Stephen D, and Joyner AL (2011). Temporal-spatial changes in Sonic Hedgehog expression and signaling reveal different potentials of ventral mesencephalic progenitors to populate distinct ventral midbrain nuclei. Neural Develop. 6, 29.

Boyden ES, Zhang F, Bamberg E, Nagel G, and Deisseroth K (2005). Millisecond-timescale, genetically targeted optical control of neural activity. Nat. Neurosci 8, 1263–1268. [PubMed: 16116447]

Bromberg-Martin ES, Matsumoto M, and Hikosaka O (2010). Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 68, 815–834. [PubMed: 21144997]

Brown J, Bullock D, and Grossberg S (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. J. Neurosci. Off. J. Soc. Neurosci 19, 10502–10511.

Burton MJ, Rolls ET, and Mora F (1976). Effects of hunger on the responses of neurons in the lateral hypothalamus to the sight and taste of food. Exp. Neurol 51, 668–677. [PubMed: 819286]

Bush RR, and Mosteller F (1951). A mathematical model for simple learning. Psychol. Rev 58, 313–323. [PubMed: 14883244]

Cardin JA, Palmer LA, and Contreras D (2008). Cellular mechanisms underlying stimulus-dependent gain modulation in primary visual cortex neurons in vivo. Neuron 59, 150–160. [PubMed: 18614036]

Chance FS, Abbott LF, and Reyes AD (2002). Gain modulation from background synaptic input. Neuron 35, 773–782. [PubMed: 12194875]

Chang CY, Esber GR, Marrero-Garcia Y, Yau H-J, Bonci A, and Schoenbaum G (2016). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat. Neurosci 19, 111–116. [PubMed: 26642092]

Cohen JY, Haesler S, Vong L, Lowell BB, and Uchida N (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482, 85–88. [PubMed: 22258508]

Contreras-Vidal JL, and Schultz W (1999). A predictive reinforcement model of dopamine neurons for learning approach behavior. J. Comput. Neurosci 6, 191–214. [PubMed: 10406133]

Danjo T, Yoshimi K, Funabiki K, Yawata S, and Nakanishi S (2014). Aversive behavior induced by optogenetic inactivation of ventral tegmental area dopamine neurons is mediated by dopamine D2 receptors in the nucleus accumbens. Proc. Natl. Acad. Sci. U. S. A 111, 6455–6460. [PubMed: 24737889]

D'Ardenne K, McClure SM, Nystrom LE, and Cohen JD (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319, 1264–1267. [PubMed: 18309087]

Day JJ, Roitman MF, Wightman RM, and Carelli RM (2007). Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nat. Neurosci 10, 1020–1028. [PubMed: 17603481]

Enomoto K, Matsumoto N, Nakai S, Satoh T, Sato TK, Ueda Y, Inokawa H, Haruno M, and Kimura M (2011). Dopamine neurons learn to encode the long-term value of multiple future rewards. Proc. Natl. Acad. Sci. U. S. A 108, 15462–15467. [PubMed: 21896766]

Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, and Uchida N (2015). Arithmetic and local circuitry underlying dopamine prediction errors. Nature 525, 243–246. [PubMed: 26322583]

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

Feierstein CE, Quirk MC, Uchida N, Sosulski DL, and Mainen ZF (2006). Representation of spatial goals in rat orbitofrontal cortex. Neuron 51, 495–507. [PubMed: 16908414]

Fiorillo CD (2013). Two dimensions of value: dopamine neurons represent reward but not aversiveness. Science 341, 546–549. [PubMed: 23908236]

Fiorillo CD, Tobler PN, and Schultz W (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299, 1898–1902. [PubMed: 12649484]

Fiorillo CD, Newsome WT, and Schultz W (2008). The temporal precision of reward prediction in dopamine neurons. Nat. Neurosci

Fiorillo CD, Yun SR, and Song MR (2013). Diversity and homogeneity in responses of midbrain dopamine neurons. J. Neurosci. Off. J. Soc. Neurosci 33, 4693–4709.

Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PEM, and Akil H (2011). A selective role for dopamine in stimulus-reward learning. Nature 469, 53–57. [PubMed: 21150898]

Friston K (2012). Prediction, perception and agency. Int. J. Psychophysiol 83, 248–252. [PubMed: 22178504]

Geisler S, and Zahm DS (2005). Afferents of the ventral tegmental area in the rat-anatomical substratum for integrative functions. J. Comp. Neurol 490, 270–294. [PubMed: 16082674]

Geisler S, Derst C, Veh RW, and Zahm DS (2007). Glutamatergic afferents of the ventral tegmental area in the rat. J. Neurosci. Off. J. Soc. Neurosci 27, 5730–5743.

Gershman SJ (2014). Dopamine ramps are a consequence of reward prediction errors. Neural Comput 26, 467–471. [PubMed: 24320851]

Glimcher PW (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc. Natl. Acad. Sci. U. S. A 108 Suppl 3, 15647–15654. [PubMed: 21389268]

Grace AA, and Bunney BS (1983). Intracellular and extracellular electrophysiology of nigral dopaminergic neurons−-1. Identification and characterization. Neuroscience 10, 301–315. [PubMed: 6633863]

Haber SN, Ryoo H, Cox C, and Lu W (1995). Subsets of midbrain dopaminergic neurons in monkeys are distinguished by different levels of mRNA for the dopamine transporter: comparison with the mRNA for the D2 receptor, tyrosine hydroxylase and calbindin immunoreactivity. J. Comp. Neurol 362, 400–410. [PubMed: 8576447]

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

Hart AS, Rutledge RB, Glimcher PW, and Phillips PEM (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. J. Neurosci. Off. J. Soc. Neurosci 34, 698–704.

Hazy TE, Frank MJ, and O'Reilly RC (2010). Neural mechanisms of acquired phasic dopamine responses in learning. Neurosci. Biobehav. Rev 34, 701–720. [PubMed: 19944716]

Hikosaka O, and Sakamoto M (1986). Neural activities in the monkey basal ganglia related to attention, memory and anticipation. Brain Dev 8, 454–461. [PubMed: 3799915]

Hollerman JR, Tremblay L, and Schultz W (1998). Influence of reward expectation on behavior-related neuronal activity in primate striatum. J. Neurophysiol 80, 947–963. [PubMed: 9705481]

Holt GR, and Koch C (1997). Shunting inhibition does not have a divisive effect on firing rates. Neural Comput 9, 1001–1013. [PubMed: 9188191]

Hong S, Jhou TC, Smith M, Saleem KS, and Hikosaka O (2011). Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. J. Neurosci. Off. J. Soc. Neurosci 31, 11457–11471.

Horvitz JC (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience 96, 651–656. [PubMed: 10727783]

Houk JC, and Davis JL (1995). Models Of Information Processing In The Basal Ganglia (MIT Press).

Howe MW, and Dombeck DA (2016). Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature 535, 505–510. [PubMed: 27398617]

Howe MW, Tierney PL, Sandberg SG, Phillips PEM, and Graybiel AM (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature 500, 575–579. [PubMed: 23913271]

Jhou TC, Fields HL, Baxter MG, Saper CB, and Holland PC (2009). The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. Neuron 61, 786–800. [PubMed: 19285474]

Jin X, and Costa RM (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature 466, 457–462. [PubMed: 20651684]

Joel D, Niv Y, and Ruppin E (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw. Off. J. Int. Neural Netw. Soc 15, 535–547.

Joshua M, Adler A, Mitelman R, Vaadia E, and Bergman H (2008). Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. J. Neurosci. Off. J. Soc. Neurosci 28, 11673–11684.

Joshua M, Adler A, Prut Y, Vaadia E, Wickens JR, and Bergman H (2009). Synchronization of midbrain dopaminergic neurons is enhanced by rewarding events. Neuron 62, 695–704. [PubMed: 19524528]

Kamin L (1969). Selective association and conditioning. In Fundamental Issues in Associative Learning, pp. 42–64.

Kato HK, Gillet SN, Peters AJ, Isaacson JS, and Komiyama T (2013). Parvalbumin-expressing interneurons linearly control olfactory bulb output. Neuron 80, 1218–1231. [PubMed: 24239124]

Kawato M, and Samejima K (2007). Efficient reinforcement learning: computational theories, neuroscience and robotics. Curr. Opin. Neurobiol 17, 205–212. [PubMed: 17374483]

Kim HF, Ghazizadeh A, and Hikosaka O (2015). Dopamine Neurons Encoding Long-Term Memory of Object Value for Habitual Behavior. Cell 163, 1165–1175. [PubMed: 26590420]

Kim Y, Wood J, and Moghaddam B (2012). Coordinated activity of ventral tegmental neurons adapts to appetitive and aversive learning. PloS One 7, e29766. [PubMed: 22238652]

Kobayashi S, and Schultz W (2008). Influence of reward delays on responses of dopamine neurons. J. Neurosci. Off. J. Soc. Neurosci 28, 7837–7846.

Kobayashi S, and Schultz W (2014). Reward contexts extend dopamine signals to unrewarded stimuli. Curr. Biol. CB 24, 56–62. [PubMed: 24332545]

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

Lammel S, Hetzel A, Häckel O, Jones I, Liss B, and Roeper J (2008). Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. Neuron 57, 760–773. [PubMed: 18341995]

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

LeCun Y, Bengio Y, and Hinton G (2015). Deep learning. Nature 521, 436–444. [PubMed: 26017442]

Lee S-H, Kwan AC, Zhang S, Phoumthipphavong V, Flannery JG, Masmanidis SC, Taniguchi H, Huang ZJ, Zhang F, Boyden ES, et al. (2012). Activation of specific interneurons improves V1 feature selectivity and visual perception. Nature 488, 379–383. [PubMed: 22878719]

Lerner TN, Shilyansky C, Davidson TJ, Evans KE, Beier KT, Zalocusky KA, Crow AK, Malenka RC, Luo L, Tomer R, et al. (2015). Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. Cell 162, 635–647. [PubMed: 26232229]

Lima SQ, Hromádka T, Znamenskiy P, and Zador AM (2009). PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. PloS One 4, e6099. [PubMed: 19584920]

Margolis EB, Lock H, Hjelmstad GO, and Fields HL (2006). The ventral tegmental area revisited: is there an electrophysiological marker for dopaminergic neurons? J. Physiol 577, 907–924. [PubMed: 16959856]

Martin SJ, Grimwood PD, and Morris RGM (2000). Synaptic Plasticity and Memory: An Evaluation of the Hypothesis. Annu. Rev. Neurosci 23, 649–711. [PubMed: 10845078]

Matsumoto M, and Hikosaka O (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447, 1111–1115. [PubMed: 17522629]

Matsumoto M, and Hikosaka O (2009a). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459, 837–841. [PubMed: 19448610]

Matsumoto M, and Hikosaka O (2009b). Representation of negative motivational value in the primate lateral habenula. Nat. Neurosci 12, 77–84. [PubMed: 19043410]

Matsumoto H, Tian J, Uchida N, and Watabe-Uchida M (2016). Midbrain dopamine neurons signal aversion in a reward-context-dependent manner. eLife 5.

Menegas W, Bergan JF, Ogawa SK, Isogai Y, Umadevi Venkataraju K, Osten P, Uchida N, and Watabe-Uchida M (2015). Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. eLife 4, e10032. [PubMed: 26322384]

Mirenowicz J, and Schultz W (1994). Importance of unpredictability for reward responses in primate dopamine neurons. J. Neurophysiol 72, 1024–1027. [PubMed: 7983508]

Miyamichi K, Shlomai-Fuchs Y, Shu M, Weissbourd BC, Luo L, and Mizrahi A (2013). Dissecting local circuits: parvalbumin interneurons underlie broad feedback control of olfactory bulb output. Neuron 80, 1232–1245. [PubMed: 24239125]

Montague PR, Dayan P, and Sejnowski TJ (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. Off. J. Soc. Neurosci 16, 1936–1947.

Morita K, Morishima M, Sakai K, and Kawaguchi Y (2013). Dopaminergic control of motivation and reinforcement learning: a closed-circuit account for reward-oriented behavior. J. Neurosci. Off. J. Soc. Neurosci 33, 8866–8890.

Morris G, Arkadir D, Nevet A, Vaadia E, and Bergman H (2004). Coincident but Distinct Messages of Midbrain Dopamine and Striatal Tonically Active Neurons. Neuron 43, 133–143. [PubMed: 15233923]

Murphy BK, and Miller KD (2003). Multiplicative gain changes are induced by excitation or inhibition alone. J. Neurosci. Off. J. Soc. Neurosci 23, 10040–10051.

Neuhoff H, Neu A, Liss B, and Roeper J (2002). I(h) channels contribute to the different functional properties of identified dopaminergic subpopulations in the midbrain. J. Neurosci. Off. J. Soc. Neurosci 22, 1290–1302.

Oleson EB, Gentry RN, Chioma VC, and Cheer JF (2012). Subsecond dopamine release in the nucleus accumbens predicts conditioned punishment and its successful avoidance. J. Neurosci. Off. J. Soc. Neurosci 32, 14804–14808.

Olsen SR, Bhandawat V, and Wilson RI (2010). Divisive normalization in olfactory population codes. Neuron 66, 287–299. [PubMed: 20435004]

Ono T, Sasaki K, Nishino H, Fukuda M, and Shibata R (1986). Feeding and diurnal related activity of lateral hypothalamic neurons in freely behaving rats. Brain Res 373, 92–102. [PubMed: 3719319]

O'Reilly RC, Frank MJ, Hazy TE, and Watz B (2007). PVLV: the primary value and learned value Pavlovian learning algorithm. Behav. Neurosci 121, 31–49. [PubMed: 17324049]

Papadopoulou M, Cassenaer S, Nowotny T, and Laurent G (2011). Normalization for sparse encoding of odors by a wide-field interneuron. Science 332, 721–725. [PubMed: 21551062]

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

Rao RP, and Ballard DH (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci 2, 79–87. [PubMed: 10195184]

Rescorla RA, and Wagner AR (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. In Classical Conditioning II: Current Research and Theory, Black A, and Prokasy W, eds. pp. 64–99.

Roesch MR, Calu DJ, and Schoenbaum G (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat. Neurosci 10, 1615–1624. [PubMed: 18026098]

Roitman MF, Wheeler RA, Wightman RM, and Carelli RM (2008). Real-time chemical responses in the nucleus accumbens differentiate rewarding and aversive stimuli. Nat. Neurosci 11, 1376–1377. [PubMed: 18978779]

Schultz W (1986). Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. J. Neurophysiol 56, 1439–1461. [PubMed: 3794777]

Schultz W (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol 80, 1–27. [PubMed: 9658025]

Schultz W (2013). Updating dopamine reward signals. Curr. Opin. Neurobiol 23, 229–238. [PubMed: 23267662]

Schultz W (2016a). Dopamine reward prediction error coding. Dialogues Clin. Neurosci 18, 23–32. [PubMed: 27069377]

Schultz W (2016b). Dopamine reward prediction-error signalling: a two-component response. Nat. Rev. Neurosci 17, 183–195. [PubMed: 26865020]

Schultz W, Apicella P, Ljungberg T, Romo R, and Scarnati E (1993). Reward-related activity in the monkey striatum and substantia nigra. Prog. Brain Res 99, 227–235. [PubMed: 8108550]

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

Sesack SR, and Grace AA (2010). Cortico-Basal Ganglia reward network: microcircuitry. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol 35, 27–47.

Shu Y, Hasenstaub A, Badoual M, Bal T, and McCormick DA (2003). Barrages of synaptic activity control the gain and sensitivity of cortical neurons. J. Neurosci. Off. J. Soc. Neurosci 23, 10388–10401.

Silver RA (2010). Neuronal arithmetic. Nat. Rev. Neurosci 11, 474–489. [PubMed: 20531421]

Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489. [PubMed: 26819042]

Stauffer WR, Lak A, Yang A, Borel M, Paulsen O, Boyden ES, and Schultz W (2016). Dopamine Neuron-Specific Optogenetic Stimulation in Rhesus Macaques. Cell 166, 1564–1571.e6. [PubMed: 27610576]

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, and Janak PH (2013). A causal link between prediction errors, dopamine neurons and learning. Nat. Neurosci

Stephenson-Jones M, Yu K, Ahrens S, Tucciarone JM, van Huijstee AN, Mejia LA, Penzo MA, Tai L-H, Wilbrecht L, and Li B (2016). A basal ganglia circuit for evaluating action outcomes. Nature 539, 289–293. [PubMed: 27652894]

Stuber GD, Klanker M, de Ridder B, Bowers MS, Joosten RN, Feenstra MG, and Bonci A (2008). Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. Science 321, 1690–1692. [PubMed: 18802002]

Sutton RS, and Barto AG (1998). Reinforcement learning: An introduction (Cambridge Univ Press).

Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, and Schoenbaum G (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. Nat. Neurosci 14, 1590–1597. [PubMed: 22037501]

Takahashi YK, Langdon AJ, Niv Y, and Schoenbaum G (2016). Temporal Specificity of Reward Prediction Errors Signaled by Putative Dopamine Neurons in Rat VTA Depends on Ventral Striatum. Neuron 91, 182–193. [PubMed: 27292535]

Tan CO, and Bullock D (2008). A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. J. Neurosci. Off. J. Soc. Neurosci 28, 10062–10074.

Tan KR, Yvon C, Turiault M, Mirzabekov JJ, Doehner J, Labouèbe G, Deisseroth K, Tye KM, and Lüscher C (2012). GABA Neurons of the VTA Drive Conditioned Place Aversion. Neuron 73, 1173–1183. [PubMed: 22445344]

Tian J, and Uchida N (2015). Habenula Lesions Reveal that Multiple Mechanisms Underlie Dopamine Prediction Errors. Neuron 87, 1304–1316. [PubMed: 26365765]

Tian J, Huang R, Cohen JY, Osakada F, Kobak D, Machens CK, Callaway EM, Uchida N, and Watabe-Uchida M (2016). Distributed and Mixed Information in Monosynaptic Inputs to Dopamine Neurons. Neuron 91, 1374–1389. [PubMed: 27618675]

Tobler PN, Fiorillo CD, and Schultz W (2005). Adaptive coding of reward value by dopamine neurons. Science 307, 1642–1645. [PubMed: 15761155]

Tsai H-C, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, and Deisseroth K (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 324, 1080–1084. [PubMed: 19389999]

Uchida N, Eshel N, and Watabe-Uchida M (2013). Division of labor for division: inhibitory interneurons with different spatial landscapes in the olfactory system. Neuron 80, 1106–1109. [PubMed: 24314722]

Ungless MA, and Grace AA (2012). Are you or aren't you? Challenges associated with physiologically identifying dopamine neurons. Trends Neurosci 35, 422–430. [PubMed: 22459161]

Vandecasteele M, Glowinski J, and Venance L (2005). Electrical synapses between dopaminergic neurons of the substantia nigra pars compacta. J. Neurosci. Off. J. Soc. Neurosci 25, 291–298.

Vitay J, and Hamker FH (2014). Timing and expectation of reward: a neuro-computational model of the afferents to the ventral tegmental area. Front. Neurorobotics 8, 4.

Volman SF, Lammel S, Margolis EB, Kim Y, Richard JM, Roitman MF, and Lobo MK (2013). New insights into the specificity and plasticity of reward and aversion encoding in the mesolimbic system. J. Neurosci. Off. J. Soc. Neurosci 33, 17569–17576.

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

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

Wenzel JM, Rauscher NA, Cheer JF, and Oleson EB (2015). A role for phasic dopamine release within the nucleus accumbens in encoding aversion: a review of the neurochemical literature. ACS Chem. Neurosci 6, 16–26. [PubMed: 25491156]

Wickersham IR, Lyon DC, Barnard RJO, Mori T, Finke S, Conzelmann K-K, Young JAT, and Callaway EM (2007). Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. Neuron 53, 639–647. [PubMed: 17329205]

Williford T, and Maunsell JHR (2006). Effects of spatial attention on contrast response functions in macaque area V4. J. Neurophysiol 96, 40–54. [PubMed: 16772516]

Wilson NR, Runyan CA, Wang FL, and Sur M (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. Nature 488, 343–348. [PubMed: 22878717]

Wise RA (2004). Dopamine, learning and motivation. Nat. Rev. Neurosci 5, 483–494. [PubMed: 15152198]

Wise RA, and Rompre PP (1989). Brain dopamine and reward. Annu. Rev. Psychol 40, 191–225. [PubMed: 2648975]

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

van Zessen R, Phillips JL, Budygin EA, and Stuber GD (2012). Activation of VTA GABA Neurons Disrupts Reward Consumption. Neuron 73, 1184–1194. [PubMed: 22445345]

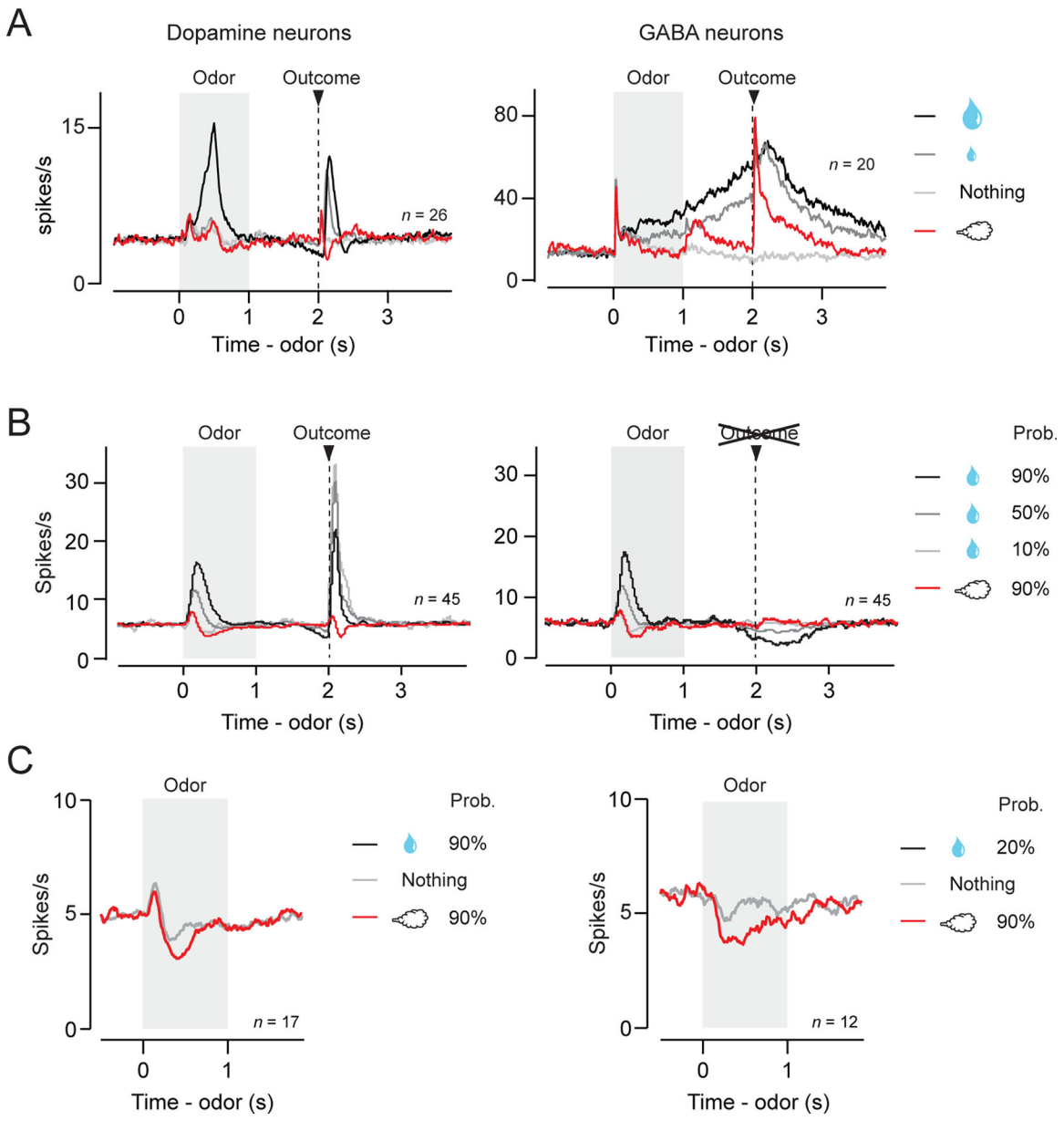

Figure 1. Firing patterns of identified dopamine and GABA neurons in VTA.

A. VTA neurons were recorded while mice performed an odor-outcome association task in which different odors predicted different outcomes (see legend on right). Odors were presented for 1 second and outcomes were presented after a 1-second delay. Neuron types were identified based on their optogenetic responses. Dopamine neurons (left, n = 26) showed phasic excitations to reward-predictive cues and reward. GABA neurons (right, n = 20) showed sustained activation during the delay. Data from Cohen et al. (2012).

B. Reward expectation modulates dopamine neuron firing. Left, when outcome was presented; Right, when outcome was omitted. Different odors predicted reward with different probabilities. Higher reward probability increased cue responses but suppressed reward responses. Data from Tian and Uchida (2015). Also see Fiorillo et al. (2003) and Matsumoto and Hikosaka (2009).

C. Reward context-dependent modulation of dopamine responses to air puff-predictive cues. The task conditions during recording differed only in the probability of reward. Dopamine neurons showed both excitation and inhibition in high-reward contexts (left) but only inhibition in low-reward contexts (right). The response in reward trials (black line) is omitted. Data from Matsumoto et al. (2016).

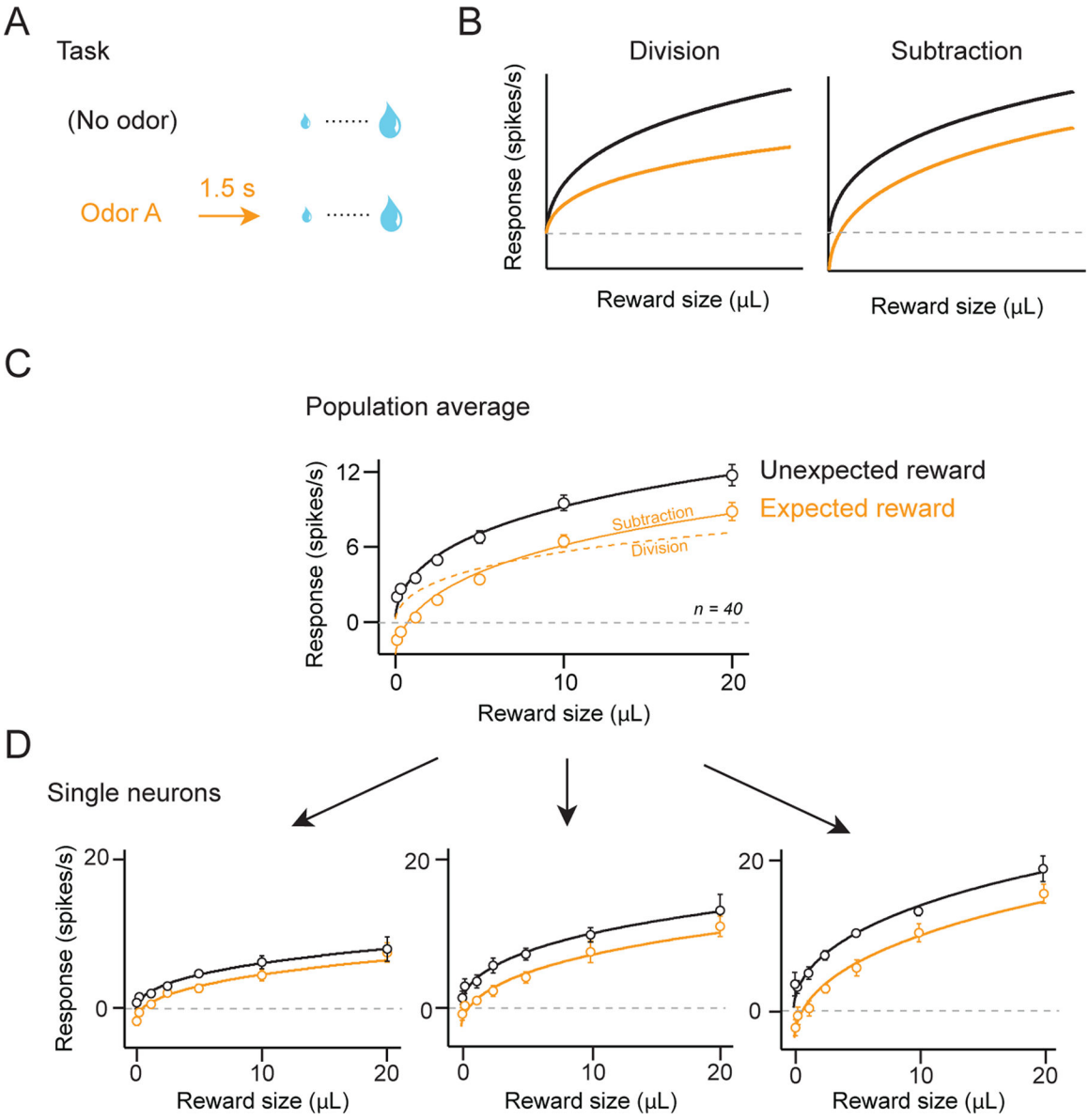

Figure 2. Subtractive computation in dopamine neurons.

A. In one task condition (No odor, black), different amounts of reward were presented without any predictive cue. In another condition (Odor A, orange), the timing of reward was predicted by an odor.

B. Prediction. Division should change the slope of the curve, whereas subtraction should cause a downward shift.

C. Average response of 40 optogenetically-identified dopamine neurons. Prediction caused a subtractive shift. Data from Eshel et al. (2015).

D. Three example neurons. Although individual neurons exhibited diversity with respect to response magnitudes, their response functions were scaled versions of one another. Data from Eshel et al. (2016).

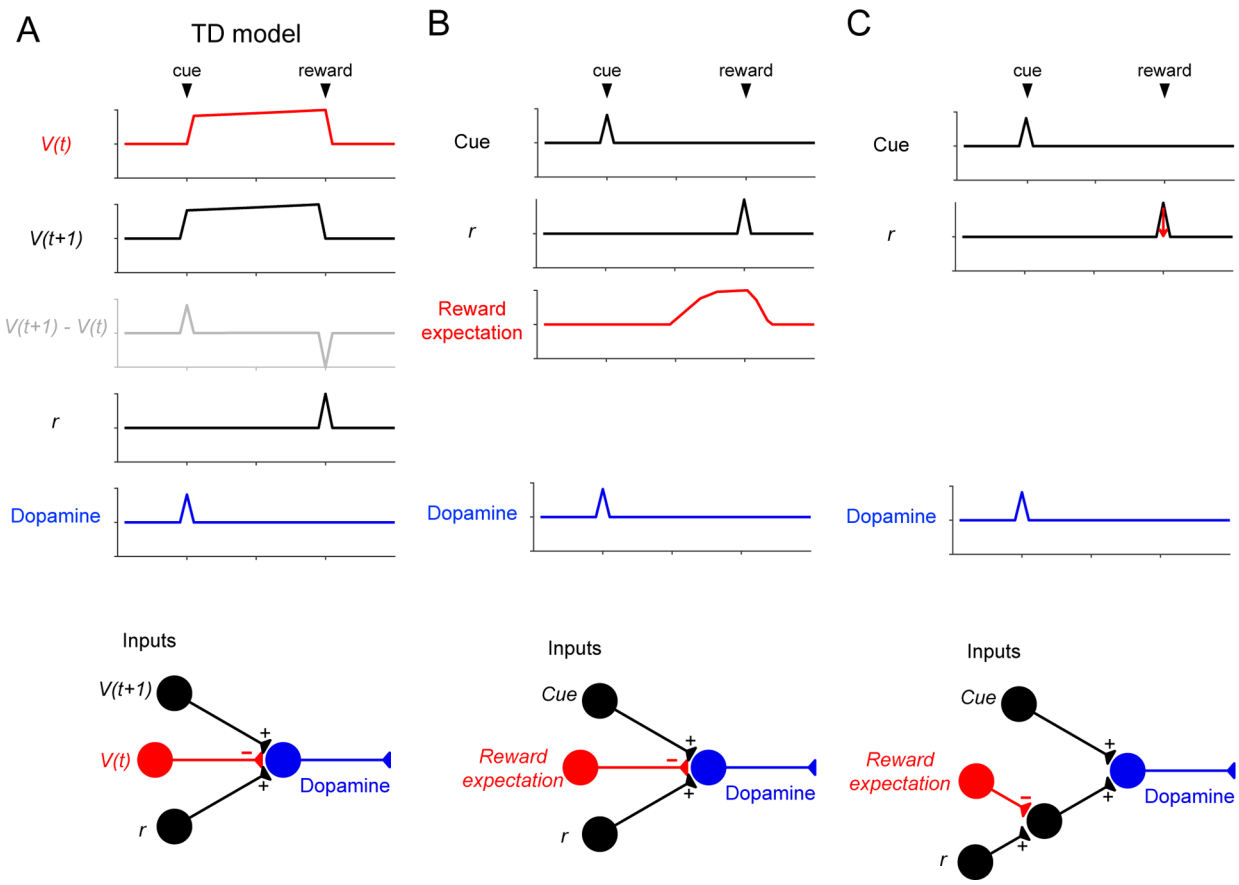

Figure 3. Models of RPE computations.

A. Temporal difference (TD) error model as implemented in Schultz et al. (1997). The computation of TD errors, $\delta = r + V(t+1) - V(t)$, can be seen as combining three inputs, one for each term. The traces show how each term changes as a function of time in a classical conditioning paradigm. The gray trace, $V(t+1) - V(t)$, can be seen as the temporal derivative of the value function, $V(t)$. The dopamine response during reward omission can be approximated by $V(t+1) - V(t)$ (gray). r: reward.

B, **C**. Alternative models assuming that reward-predictive cues and reward elicit phasic excitation. Reward expectation modulates dopamine reward responses either at the dopamine neuron itself (**B**) or upstream (**C**).

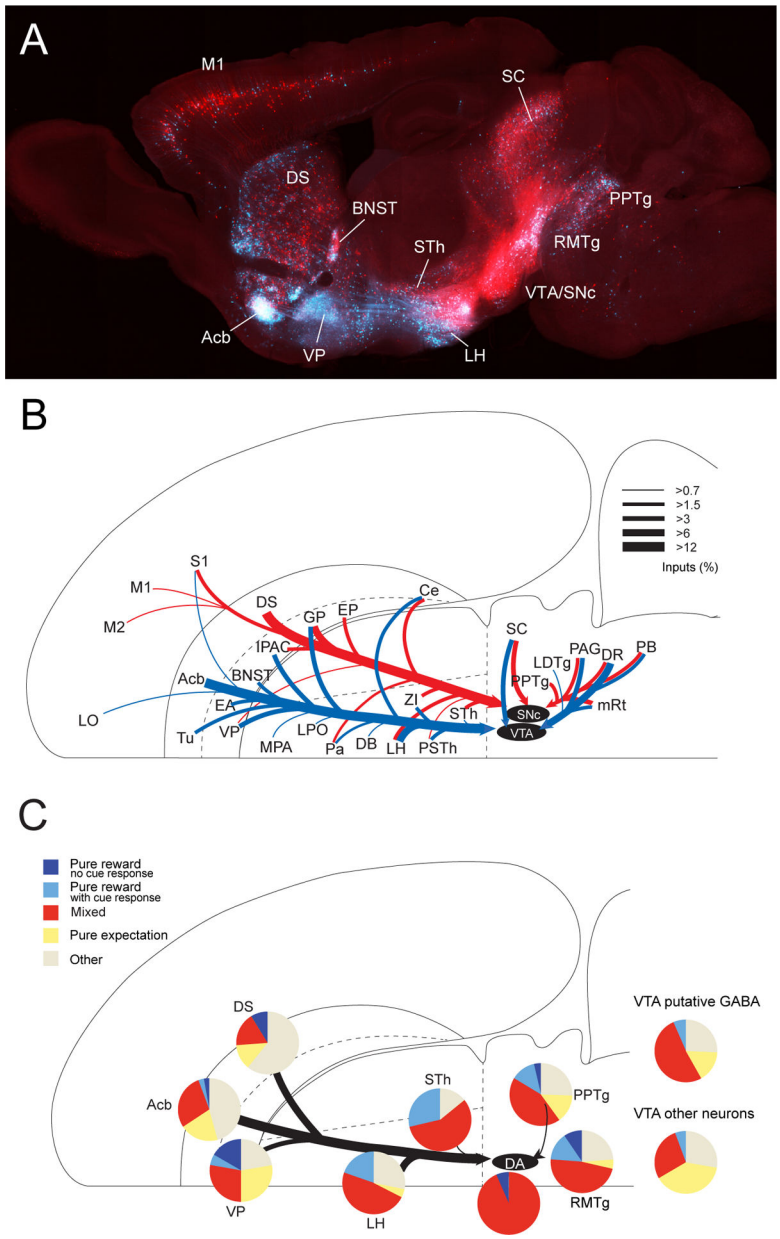

Figure 4. Monosynaptic input to dopamine neurons.

A. Monosynaptic inputs to VTA and SNc dopamine neurons (blue and red, respectively). Inputs were labeled through transsynaptic retrograde tracing using rabies virus. Data from Watabe-Uchida et al. (2012).

B. Schematic summary of A. The thickness of each line indicates the extent of inputs from each area (% of total inputs).

C. Firing patterns of monosynaptic inputs in a classical conditioning paradigm. Monosynaptic inputs to dopamine neurons were labeled by channelrhodopsin-2 using rabies virus. Optogenetics were used to identify these inputs in 7 brain areas while mice performed a task. Data from Tian et al. (2016).

LO: lateral orbitofrontal cortex; M1, primary motor cortex; M2, secondary motor cortex; S1, primary somatosensory cortex; Tu, olfactory tubercle; Acb, nucleus accumbens; DS, dorsal striatum; VP, ventral pallidum; EA, extended amygdala; BNST, bed nucleus of stria terminalis; IPAC, interstitial nucleus of the posterior limb of the anterior commissure; GP, globus pallidus (external segment of the globus pallidus); EP, entopeduncular nucleus (internal segment of the globus pallidus); MPA, medial preoptic area; LPO, lateral preoptic area; Pa, paraventricular hypothalamic nucleus; DB, diagonal band of Broca; Ce, central amygdala; LH, lateral hypothalamus; ZI, zona incerta; STh, subthalamic nucleus; PSTh, parasubthalamic nucleus; SC, superior colliculus; PPTg, pedunculopontine tegmental nucleus; LDTg, Laterodorsal tegmental nucleus; PAG, Periaqueductal gray; DR, dorsal raphe; mRt, Reticular formation; PB, parabrachial nucleus.