Neural Circuit Mechanisms Underlying Dopamine Reward Prediction Errors

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Tian, Ju. 2016. Neural Circuit Mechanisms Underlying Dopamine Reward Prediction Errors. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493547">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493547</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Neural circuit mechanisms underlying dopamine reward prediction errors

A dissertation presented
by
Ju Tian
to
The Division of Medical Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Neurobiology

Harvard University
Cambridge, Massachusetts
April 2016
Neural circuit mechanisms underlying dopamine reward prediction errors

Abstract

Dopamine neurons are thought to facilitate learning by signaling reward prediction errors (RPEs), the discrepancy between actual and expected reward. However, how RPEs are calculated remains unknown. In Chapter 1, I tested the hypothesis that RPE signals in dopamine neurons are inherited entirely from the lateral habenula, by examining how lesions of the habenular complex affect the response of optogenetically-identified dopamine neurons in mice. I found that despite large lesions of habenula, dopamine neurons maintained features of RPE coding pertaining to phasic dopamine responses. Interestingly, a specific aspect of RPE signaling—the inhibitory responses caused by reward omission—was greatly diminished in habenula lesion animals. These results suggested that the RPE signals in dopamine neurons were not simply relayed from habenula and that multiple mechanisms underlie RPE signaling.

In Chapter 2, I systematically examined how RPE is computed at a neural circuit level, by combining rabies virus-based monosynaptic retrograde tracing with optogenetic cell identification during electrophysiological recording. We characterized the firing patterns of 205 neurons presynaptic to dopamine neurons (“input neurons”) from 7 major VTA input areas in behaving mice. We found that relatively few input neurons signaled purely ‘actual’ reward or ‘expected’ reward. Instead, many input neurons across brain areas signaled combinations of these types of information. We also found that some input neurons signaled already-computed RPEs. These results demonstrate that the information required for dopamine neurons to compute...
RPE is not localized to specific brain areas; rather, the computation is distributed across multiple nodes in a brain-wide network. Together, these results provide new insights on the neural circuits involved in the computation of RPE signals in dopamine neurons.
# Table of Contents

Introduction........................................................................................................................................1

Chapter 1. Habenula Lesions Reveal that Multiple Mechanisms Underlie Dopamine Prediction Errors................................................................................................................23

Chapter 2. Distributed and multi-step computation of dopamine prediction errors ........63

Conclusion........................................................................................................................................110

References......................................................................................................................................122
Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>Dopamine signals resemble temporal difference errors</td>
<td>8</td>
</tr>
<tr>
<td>0.2</td>
<td>A modified rabies virus system labels the monosynaptic inputs of dopamine</td>
<td>19</td>
</tr>
<tr>
<td>1.1</td>
<td>Odor-outcome association task</td>
<td>26</td>
</tr>
<tr>
<td>1.2</td>
<td>Electrolytic lesions in habenula and recording sites in VTA region</td>
<td>27</td>
</tr>
<tr>
<td>1.3</td>
<td>Habenula lesion altered the anticipatory licking behavior</td>
<td>29</td>
</tr>
<tr>
<td>1.4</td>
<td>Optogenetic identification of dopamine neurons</td>
<td>30</td>
</tr>
<tr>
<td>1.5</td>
<td>The overall firing patterns of dopamine neurons</td>
<td>32</td>
</tr>
<tr>
<td>1.6</td>
<td>Inhibitory responses during reward omission were diminished in lesion animals</td>
<td>34</td>
</tr>
<tr>
<td>1.7</td>
<td>Dopamine neurons’ response in pre-reward period</td>
<td>35</td>
</tr>
<tr>
<td>1.8</td>
<td>Inhibitory responses to air puff-predictive cues and air puff were unimpaired.</td>
<td>37</td>
</tr>
<tr>
<td>1.9</td>
<td>Prediction error signals in air puff response</td>
<td>38</td>
</tr>
<tr>
<td>1.10</td>
<td>Response to air puff in different time windows</td>
<td>39</td>
</tr>
<tr>
<td>1.11</td>
<td>Phasic excitations to reward CS and US were weakened in lesion animals</td>
<td>42</td>
</tr>
<tr>
<td>1.12</td>
<td>Dopamine neurons’ response from animals with large lesions</td>
<td>43</td>
</tr>
<tr>
<td>1.13</td>
<td>Comparison of all VTA neurons’ responses in control and lesion group</td>
<td>45</td>
</tr>
<tr>
<td>1.14</td>
<td>Simulation of anticipatory licking based on the Rescorla-Wagner model</td>
<td>48</td>
</tr>
<tr>
<td>2.1</td>
<td>Experimental design and behavioral task</td>
<td>68</td>
</tr>
<tr>
<td>2.2</td>
<td>Recording of VTA neurons in rabies virus-infected or AAV-infected control</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>animals</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Comparison of RPE signals between rabies infected and AAV infected dopamine</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>neurons</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>Histological verification of the recording locations of input neurons</td>
<td>72</td>
</tr>
<tr>
<td>2.5</td>
<td>Each input area contains input neurons with diverse response profiles</td>
<td>74</td>
</tr>
</tbody>
</table>
Figure 2.6 Response profiles of input neurons in trials with probabilistic rewards. .......................... 76
Figure 2.7 Correlation analysis of dopamine neurons and their inputs............................................ 77
Figure 2.8 Mixed coding of reward and expectation signals in input neurons................................. 79
Figure 2.9 Criteria used to categorize neurons into pure reward and pure expectation types......... 82
Figure 2.10 Input neurons encode reward prediction error signals................................................. 84
Figure 2.11 Criteria used to categorize neurons into RPE-related response types.......................... 85
Figure 2.12 Light identification criteria and rabies recording days do not affect the main results.. 87
Figure 2.13 Mixed and RPE types of neurons are more concentrated in other (non-light identified) neurons........................................................................................................................................... 88
Figure 2.14 Comparison of CS responses in identified input neurons and other neurons.......... 90
Figure 2.15 Comparison of CS responses in identified input neurons and other neurons.......... 91
Figure 2.16 A Linear model to fit dopamine neurons’ task responses using inputs. ...................... 92
Figure 2.17. A linear model can reliably extract RPE signals from inputs. .................................... 93

Figure 3.1 Lateral habenula projects to input areas of dopamine neurons.................................113
Acknowledgements

First, I would like to thank my PhD advisor, Nao. I’m very grateful that Nao gave me the opportunity to work in his lab, even though I had little experience in the field. Nao has provided me a lot of freedom to explore my interests. Whenever I need help, he is always ready to provide the critical inputs to drive projects forward. Nao is also an excellent listener. When I encounter failures and frustrations with experiments, Nao’s words always fill me with courage. Without Nao’s encouragements and guidance, I cannot finish the work in this dissertation.

I also want to thank all of the other colleagues in the lab. Thanks for being helpful throughout the years and making the lab such a fun place to work! Especially, I would like to thank Mitsuko, who has worked together with me on the recording project using rabies virus. Without her pioneer work in putting together all of the techniques, and her courage of even attempting to try it, this project would not be possible at all. Her persistent pursuit of truth in science and her curiosity about nature are always inspiring to me. Jeremiah and Sebastian advised me during my rotation. They taught me many of the experimental skills I need to finish this dissertation. Moreover, Jeremiah also taught me a lot on how to be a good lab member. I want to thank all of the undergrads who had worked with me: Lena, Marissa and Ryan. Although I was the mentor, I learned so much from you.

I would like to thank my dissertation advisory committee: Bernardo Sabatini, Mark Andermann, Bence Olveczky and John Maunsell. Thanks for providing me valuable advice during all of the committee meetings and keeping my project on track.

I also want to thank my exam committee: Peter Dayan, Stefano Fusi, John Assad, and the chair, Mark Andermann. Thanks for taking time out of your busy schedule to read my dissertation and coming to Boston for the exam. It is a great honor to have you all.

I would like to thank Ed and Joel, two very talented engineers from the Center of Brain Science, and Raji, a graduate student from Bence’s lab. Thanks you for helping me to set up the customized recording system, which has saved me a lot of time in doing recording experiments. Ed and Joel have also helped me in building many other lab equipments. Without you, my PhD could have been twice as long.

Lastly, I want to thank my family members. My parents have always been very supportive on my education and career. My brother triggered my interests in science when I was little, and has been taking care of my parents while I am abroad. My boyfriend, Raymond, has given me a lot of encouragements and advice when I was in difficult moments. I am grateful to have you in my life.
INTRODUCTION
Introduction

The human brain, as an information processing system, has extraordinary learning abilities. Learning can occur in diverse domains, from playing piano, to mastering a new language, to finding out what the best local Chinese restaurant is. The ability to assign positive and negative value to environmental stimuli is fundamental for animals’ survival. However, learning accurate value representations is not a computationally easy task for several reasons. First, we receive wealth sensory inputs at any given moment. In a high dimensional sensory space, it is difficult to robustly extract the key features that are relevant to rewards or punishments. Second, rewards and punishments often occur long after the presentation of the relevant stimuli or actions, which complicates attribution of the outcome to specific events. Lastly, because the environment is always changing, the learning agent cannot settle on an single optimal strategy. Instead, the agent has to constantly balance exploration and exploitation.

In the past fifty years, research in animal learning has identified the key elements underlying simple associative learning involving presentation of stimuli followed by reward. In parallel, the machine learning community has developed algorithms as to how an agent could maximize rewards while interacting with its environments, providing mathematical models for associative learning. Recently, neurophysiologists have attempted to reveal the neural basis of associative learning by identifying neural signals that correspond to these mathematical models. One major neurophysiological finding is that dopamine neurons in the midbrain seem to encode reward prediction error signals, which could serve as the error signals proposed in animal learning theory and some machine learning algorithms. In the following sections, I will review some of the classical works from these different fields and describe how these studies motivate my dissertation.
Classical conditioning: experiments and theories

Studying a simple behavioral paradigm such as classical conditioning is well-suited towards revealing the basic rules of animal learning behavior. In the well-known experiment, Professor Pavlov rang a bell (conditioned stimulus, CS) right before giving a food reward (unconditioned stimulus, US) to a hungry dog. After repeating this procedure many times, the dog started to salivate (conditioned response) in response to the bell sound. In other words, the dog learned that the bell sound predicted the future food. At a first glance, it seems that the temporal contiguity between bell sound and food caused the animal to associate the two. However, in a later experiment, Kamin found that temporal contiguity is not sufficient for the learning (Kamin, 1969). In this experiment, the animal was trained to associate CS B with a US. When CS A was presented together with CS B, the presence of CS B blocked the learned association between CS A and US. Kamin interpreted this effect as following: “Perhaps, for an increment in an associative connection to occur, it is necessary that the US instigate some mental work on the part of the animal. This mental work will occur only if the US is unpredicted, if it in some sense surprises the animal” (Kamin, 1969). Notably, the original experiments by Kamin used a conditioned suppression procedure, in which rats learned to stop pressing a food lever to avoid foot-shock. Later studies showed that this blocking effect also occurs in conditioning paradigms using an appetitive US (Khallad and Moore, 1996; Steinberg et al., 2013; Waelti et al., 2001), suggesting the generality of the blocking effect.

Since the discovery of blocking, many models were proposed to explain it. One of the most influential models is the Rescorla-Wagner model. The key message from the blocking experiment is that the US must be somewhat surprising to trigger learning. One intuitive way to think about surprise is to define it as how much an outcome deviates from the expected outcome.
This forms the basis of the Rescorla-Wagner Model. Here we use associative strength (V) between CS and US to indicate the extent of the association. $V = 0$ means no association has been formed at all; $V = \lambda$ is the maximum associative strength induced by the US used in training. The model assumes that learning (i.e. the change of associative strength after each trial) is proportional to the amount of surprise (i.e. the difference between US and expectation). This is formulated as Equation 1 and 2, where $\alpha_A$ is the learning rate of CS A and $\alpha_B$ is the learning rate of CS B. In the case of a compound CS, the model further assumes that the level of expectation is equal to the sum of the associative strength of all individual CSs, see Equation 3. In the blocking experiment, because CS A has already been fully associated with the US, $V_A = \lambda$. The total associative strength $V = V_A + V_B = \lambda + 0 = \lambda$. According to Rescorla-Wagner model, since the total associative strength is equal to the maximum associative strength, no learning occurs in CS B. Therefore, the Rescorla-Wagner Model elegantly explains the blocking effect.

$$\Delta V_A = \alpha_A \ast (\lambda - V)$$  \hspace{1cm} (1)
$$\Delta V_B = \alpha_B \ast (\lambda - V)$$  \hspace{1cm} (2)
$$V = V_A + V_B$$  \hspace{1cm} (3)

The Rescorla-Wagner model impacted the associative learning field greatly not only because it provided a theoretical account for many experimental observations at the time, but also because it made some interesting predictions and that later experiments showed to be true. In a two-phase learning paradigm, two different CSs are separately conditioned with the US until the learning reaches asymptote. In the second phase, the two CSs are presented as a compound cue and paired with the same US. Since the CS is always paired with US, intuitively, the associative strength should not change. However, based on the Rescorla-Wagner model, the
associative strength of each cue will be predicted to gradually decrease since the summed value
prediction of the compound cue is higher than the single US. It was later shown that this
prediction from the Rescorla-Wagner model is correct (Khallad and Moore, 1996; Kremer, 1978).

Despite the great success of Rescorla-Wagner model, this model also has some
shortcomings. An important feature that the Rescorla-Wagner model does not capture is the
temporal structure between stimuli. Going back to the blocking paradigm, we know that if CS A
has already been fully associated with the US, simultaneously presenting CS A and B would
block the learning of the association between CS B and US. Now imagine, instead of presenting
CS A and B simultaneously, we present CS B right before CS A: would the animal learn to
associate CS B and US? Interestingly, under this configuration, the associative strength of CS B
gradually increases and the associative strength of CS A gradually decreases (David and Miller,
1962). This phenomenon is called the primacy effect and this effect cannot be accounted for by
trial-based models such as the Rescorla-Wagner model.

To better account for the timing of stimuli in classical conditioning, models that can be
applied continuously, on a moment-by-moment rather than trial-by-trial basis, have been
proposed. Among these models, the most influential is the temporal difference (TD) model
(Sutton, 1988a). TD models view the conditioning process as learning to predict the summed
intensity of future USs, namely the value function: V. V has some analogy to the associative
strength in Rescorla-Wagner model: the stronger the stimulus that has been associated with the
reinforcements, the bigger the value of V is. The key difference is that, V in the Rescorla-
Wagner model is a function only of trial number; whereas in the TD model, V is a function of
states (e.g. time elapsed from trial-start). The TD model assumes that, at each moment, the
animal predicts the sum of all future reinforcements, placing higher weights on proximal reward and lower weight on delayed reward. This can be formulated as follows:

\[ V(t) = \sum_{i=0}^{\infty} \gamma^i r(t + i) \]  

(4)

In this equation, \( V(t) \) denotes the predicted future reward at time \( t \); \( \gamma \) is the discount factor with a value between 0 and 1, where smaller values more steeply decrement delayed rewards; \( r(t) \) represents the intensity of reinforcement at time \( t \). Equation 4 can be rewritten in the following way:

\[ V(t) = r(t) + \gamma \sum_{i=0}^{\infty} \gamma^i r(t + 1 + i) \]  

(5)

it should be the case that:

\[ V(t + 1) = \sum_{i=0}^{\infty} \gamma^i r(t + 1 + i) \]  

(6)

in this circumstance, equation 5 is equivalent to:

\[ V(t) = r(t) + \gamma * V(t + 1) \]  

(7)

The goal of the subject is to learn the correct \( V(t) \) by experience. Imagine a scenario in which the subject receives an unexpected reinforcement at time \( t \). This causes a discrepancy between the right and left side of Equation 7. This discrepancy \( \delta(t) \) is called the temporal difference error, formulated in Equation 8. To correct for the discrepancy and to learn the accurate \( V(t) \), the subject needs to update the value of \( V(t) \) after experiencing \( r(t) \), as shown in Equation 9. \( \alpha \) in Equation 9 is the learning rate, and bigger value of \( \alpha \) indicates faster learning rate.
\[
\delta(t) = r(t) + \gamma \cdot V(t + 1) - V(t)
\]
\[
V(t) \leftarrow V(t) + \alpha \cdot \delta(t)
\]

It can be proven mathematically that the solution of the TD model will converge to the correct value function for any given configuration of the task (Dayan, 1992; Sutton, 1988a; Sutton and Barto, 1998; Watkins and Dayan, 1992). The TD model has accurately predicted animals’ behavior in a variety of conditioning paradigms. For instance, the model recapitulates the effect that the association of a remote stimulus can be enhanced by adding an intervening stimulus. The model can also explain the aforementioned temporal primacy effect (Sutton and Barto, 1990). In summary, the computational goal of the TD model is to predict at each moment the sum of discounted sum of future reinforcements. To achieve this goal, the TD model updates the value prediction by adding a temporal difference error signal. The TD model not only predicts features of classical conditioning, but also provides a potential mechanistic explanation for how learning happens.

The role of dopamine in associative learning

As we can see in the previous section, the temporal difference error signal is critical for TD learning. If an animal’s brain is also using an algorithm similar to TD learning, can we find neural populations that represent this temporal difference error signals? Substantial evidence suggests that midbrain dopamine neurons encode an error signal that is very similar to the one proposed in the TD learning algorithm.

Wolfram Schultz and his colleagues found that putative dopamine neurons showed a peculiar set of responses when monkeys performed a classical conditioning paradigm. When a drop of juice reward was delivered unexpectedly to the animal, dopamine neurons fired a burst of
spikes. After the animal learned to associate a visual cue with upcoming reward, dopamine neurons instead fired at the cue onset rather than at the time of reward. Interestingly, if an expected reward was omitted unexpectedly, dopamine neurons briefly paused their firing. Based on these evidence, it has been postulated that dopamine neurons encoded the temporal difference error signal similar to those used in reinforcement learning theories (Figure 0.1) (Houk et al., 1995a; Schultz et al., 1997).

![TD model](image)

**Figure 0.1** Dopamine signals resemble temporal difference errors. 

$r(t)$ represents the actual reward, and $V(t)$ represents the summation of all predicted future rewards at time $t$. After the animal has fully learned the association between a cue and reward, $V(t)$ increases at the onset of the sensory cue predicting rewards and goes back to zero after reward delivery. By combining the $r(t)$, $V(t+1)$ and a negative $V(t)$, one can get TD error signals. Consistent with TD error signals, dopamine neurons are only activated by the reward-predicting cues but not by the predicted reward. For simplicity, the temporal discount factor is set to 1 in the illustration.

This error signal encoded by dopamine neurons has been shown to vary with many of the features associated with reward value, including magnitude (Bayer and Glimcher, 2005; Tobler...
et al., 2005), probability (Fiorillo, 2003; Tian and Uchida, 2015), timing (Fiorillo et al., 2008; Kobayashi and Schultz, 2008), and subjective preference (Lak et al., 2014).

Whether dopamine neurons encode prediction error accurately in the negative domain has been controversial. This has been studied both in the case of reward that is smaller than expected and of aversive stimuli. In a classic study of reward responses, Bayer et al. found that the firing rates of dopamine neurons cannot be explained by weighted sum of reward history only when dopamine responses were below baseline (Bayer and Glimcher, 2005). Later, it has been shown that dopamine neurons seem to encode negative reward prediction error by pause durations (Bayer et al., 2007). These results suggest dopamine neurons encode reward prediction error signals in a nonlinear way. On the other hand, electrochemical measurements of dopamine release in the nucleus accumbens core showed that dopamine signals encode reward prediction error faithfully in both the positive and negative directions (Hart et al., 2014). This suggests that even though the negative prediction errors are not well encoded in firing rates, they might be well encoded in the concentration of dopamine release in downstream areas. Therefore, along the axis of reward delivery, dopamine signaling seems to conform to the requirements of TD error.

However, in respect to responses to aversive stimuli, the debate still continues. It is thought that most dopamine neurons are inhibited by aversive stimuli (Mileykovskiy and Morales, 2011; Tan et al., 2012; Ungless et al., 2004). However, other studies also found that dopamine neurons can be activated by aversive stimuli, although the proportion and location of dopamine neurons reported varied across studies (Brischoux et al., 2009; Guarraci and Kapp, 1999; Joshua et al., 2008; Lammel et al., 2012; Matsumoto and Hikosaka, 2009). Furthermore, a recent study even suggests that dopamine neurons completely ignore aversive stimuli and only signal prediction errors along the reward axis (Fiorillo, 2013). The discrepancy between results might be due to
the differences in behavior tasks, types of aversive stimuli, recording locations, and the method used to identify dopamine neurons. More studies are needed to clarify the nature of dopamine responses to aversive stimuli.

If the reward prediction error signals encoded by dopamine neurons correspond to the TD error signal, dopamine activity could be utilized to guide animal learning. Indeed, different lines of evidence have suggested that dopamine activity is important for associative learning. Pharmacological blockage of dopamine receptors can reduce lever-pressing for brain stimulation reward, food reward, and also prevent the animals from developing place preference after experiencing reward (Wise, 2004).

Recent developments of optogenetic tools make it possible to probe the function of dopamine activity in a temporally precise manner. Transgenic mice (Bäckman et al., 2006) and rats (Witten et al., 2011) facilitate the expression of channelrhodopsin (Atasoy et al., 2008), a light gated cation channel, specifically in dopamine neurons. Using these techniques, it has been shown that phasic activation of dopamine neurons can induce place preference in mice (Tsai et al., 2009), or serve as a positive reinforcer for operant conditioning (Adamantidis et al., 2011; Kim et al., 2012; Pascoli et al., 2015; Rossi et al., 2013). Moreover, in the aforementioned blocking paradigm, optogenetic activation of dopamine neurons at the time when reward is delivered can unblock the learning of the new CS presented in the compound stimuli, establishing a causal relationship between dopamine activity and learning (Steinberg et al., 2013). Conversely, optogenetic inactivation of dopamine neurons by expressing halorhodopsin (Gradinaru et al., 2008), a light gated chloride pump, mimics the behavioral effect observed after an animal experiences negative prediction errors (Chang et al., 2016). In summary, the dopamine
prediction error signal is similar to the error signal critical to the TD model, and is also important for normal learning behaviors.

How do dopamine neurons affect downstream areas and promote learning? Anatomically, dopamine neurons project broadly to many brain areas, including striatum, septum, amygdala and some cortical areas (Swanson, 1982; Yetnikoff et al., 2014). The most intensively innervated downstream area is the striatum, with the dorsal striatum receiving more projections from SNc dopamine neurons and the ventral striatum receiving more projection from VTA dopamine neurons (Yetnikoff et al., 2014). It is thought that cortical inputs to the striatum carry information about behavioral states and that the striatum performs action selection based on this contextual information. It has been hypothesized that RPE signals from dopamine neurons modulate the plasticity of cortico-striatal synapses, strengthening connections that promote rewarding actions and weakening connections that lead to aversive outcomes. A large amount of in vitro studies have shown that bath application of dopamine or dopamine receptor agonists can modulate cortico-striatal synaptic transmission (Tritsch and Sabatini, 2012). Moreover, it has also been shown in vivo that self-stimulation with an electrode in the substantia nigra could induce potentiation of cortico-striatal synapses in a way that depends on activation of dopamine receptors (Reynolds et al., 2001). Nevertheless, most of these studies do not activate dopamine receptors in a way that is similar to physiological conditions, which is required to understand the physiological effects of dopamine release in downstream areas.

Other functions of dopamine

In addition to modulating neural plasticity and promote learning, dopamine is also important normal movements. Parkinson disease patients have a great loss of dopaminergic cells
(Bernheimer et al., 1973) and motor dysfunction such as tremor, bradykinesia, and loss of postural reflexes are considered core signs of the disease (Jankovic, 2008). Dopamine’ function in movements led to the discovery of dopamine as a novel neurotransmitter back in the 1950s. Before dopamine was known as an independent neurotransmitter in the nervous system, dopamine had long been known as a precursor for norepinephrine. Arvid Carlsson and his colleagues found that the akinetic effects of reserpine can be rescued by intravenous injection of 3,4-dihydroxyphenylalanine (DOPA), a precursor of dopamine and norepinephrine (Carlsson et al., 1957, 1958). Moreover, recovery from the akinetic effects was correlated with the recovery of dopamine levels, rather than norepinephrine levels, suggesting that the depletion of dopamine led to the akinetic effects. This implies dopamine may work as an independent neurotransmitter. Clinically, supplementing Parkinson disease patients with L-3,4-dihydroxyphenylalanine can effectively reduce Parkinsonian symptoms (Cotzias et al., 1967).

Dopamine can also modulate reward-driven movements. Studies from Hikosaka’s group demonstrated this in monkeys performing a visually guided saccade task (Hikosaka et al., 2014; Nakamura and Hikosaka, 2006). Monkeys were trained to saccade either left or right based on a visual cue to obtain reward. The reward sizes of left and right sides were not equal. In different blocks, the reward sizes on the left and right were reversed. Monkeys showed faster saccade reaction times when they made saccades to the large reward target than to the small reward target. When dopamine D1 receptors in the striatum were blocked pharmacologically, the difference in saccade latency between large and small reward became smaller, mainly by increasing the saccade latency to big reward target; when dopamine D2 receptors in the striatum were blocked, the bias of saccade latency by reward became larger, mainly due to an increase in the saccade
latency to small reward target. These results suggest that dopamine plays an important role in reward-dependent modulation of saccade movement.

Although the loss of dopamine neurons is related to the movement disorder in Parkinson’s patients, phasic dopamine activity does not seem to be affected by movements. Delong et al. has shown that putative dopamine neurons in the substantia nigra pars compacta (SNc) do not seem to be affected by active movements or passive manipulations (DeLong et al., 1983). Consistent with this finding, Romo and Schultz found that dopamine neurons show a very mild increase in activity before and during arm reaching movements (Romo and Schultz, 1990). These lines of evidence suggest that tonic dopamine levels might be important for normal movements, while phasic dopamine activity drives learning by signaling reward prediction errors.

How might dopamine neurons compute reward prediction error?

Although establishing dopamine neurons as a neural substrate for prediction error signals has been a great step towards understanding the mechanism underlying associative learning, many other questions remain to be answered. One important question asks how prediction error signals are generated in the brain. Addressing this question can help us understand how a reinforcement learning algorithm is implemented in the brain. Dopamine neurons receive broad inputs from many brain regions (Beier et al., 2015; Menegas et al., 2015; Watabe-Uchida et al., 2012). Based on anatomical connections and the firing patterns of neurons in these input areas, many models have been proposed to explain the mechanism underlying the generation of prediction error signals in dopamine neurons. In this section, I will review some of these models together with some recent experimental data. Dopamine neurons are mainly located in two brain areas: ventral tegmental area (VTA) and SNc. With respect to their responses to reward-related
events, dopamine neurons in both VTA and SNc encode reward prediction error (Fiorillo, 2013; Matsumoto and Hikosaka, 2009). Therefore, although we focus our discussion on VTA dopamine neurons, what we learn about VTA dopamine neurons in computing reward prediction error will likely generalize to SNc dopamine neurons.

**Local VTA circuitry**

Dopamine neurons receive strong inputs from local interneurons (Dobi et al., 2010). About 55-65% of the VTA neurons are dopaminergic, and the rest are mostly GABAergic inhibitory neurons (Nair-Roberts et al., 2008; Swanson, 1982). For a long time, it has been technically difficult to identify different cell types during in vivo recording. The conventional approach to distinguish dopamine neurons from other neurons during in vivo recordings relies on the spike waveforms and other firing properties. However, this approach has recently been called into question (Lammel et al., 2008; Margolis et al., 2006; Ungless and Grace, 2012).

To distinguish genetically distinct cell types unequivocally, Cohen and Haesler et al. used an optogenetic tagging technique taking advantage of transgenic mice that express Cre recombinase in specific cell types (Cohen et al., 2012). By expressing channelrhodopsin specifically in dopamine neurons, only dopamine neurons would fire action potentials reliably within a short latency of laser stimulation. By using this technique, the authors found that optogenetically-identified dopamine neurons in the VTA showed phasic excitations to reward as well as to cues predicting reward. These identified dopamine neurons also dipped below baseline when an expected reward was omitted. These results are consistent with recording results from primates, suggesting that the prediction error signals in dopamine neurons are well-conserved between primates and rodents.
Interestingly, Cohen and Haesler et al. also found that the response profiles of VTA neurons in a classical conditioning task fell into three clusters. Neurons in the first cluster showed phasic responses to reward predictive cues and rewards. All identified dopaminergic neurons fell into this cluster. Neurons in the second cluster showed a sustained increase of activity in trials in which a big reward was expected. All identified VTA GABA neurons fell into the second cluster. Lastly, neurons in the third cluster decreased their activities after cues predicting big reward. These results showed that different cell types in VTA also corresponded to different firing patterns: dopamine neurons encode reward prediction error signals and VTA GABA neurons encode reward expectation signals that could correspond to reward expectation.

Can VTA GABA neurons provide the expectation signals for dopamine neurons to compute reward prediction error? Optogenetic stimulation of VTA GABA neurons reduce the excitability and evoked activity of VTA dopamine neurons (van Zessen et al., 2012). Therefore, when reward is expected, the inhibitory GABAergic input could suppress the excitatory reward input and produce the reward prediction error signal in dopamine neurons. To test this hypothesis, Eshel et al. used optogenetics to selectively stimulate or inhibit VTA GABA neurons while recording from dopamine neurons (Eshel et al., 2015). The authors found that stimulating VTA GABA neurons suppressed dopamine reward responses, as if reward were expected. Furthermore, this suppression took the same arithmetic form—subtraction—as natural, cue-driven expectation. Conversely, inhibiting VTA GABA neurons increased dopamine reward responses, as if reward were less expected. Finally, bilaterally stimulating VTA GABA neurons while keeping rewards unchanged caused mice to learn new, reduced values for odor stimuli, consistent with prediction error theory. These results suggest that VTA GABA neurons play an important role in dopamine
RPE signaling. From this study, however, it remains unknown whether VTA GABA neurons are the only source of expectation signals for dopamine neurons.

**Lateral habenula as a source of prediction error signals**

While dopamine neurons could compute prediction errors locally by integrating different input signals, dopamine neurons could also simply relay the computed prediction error signals from an upstream area. Recently, lateral habenula (LHb) neurons have been shown to signal reward prediction errors in a way that is opposite to signals from dopamine neurons (Matsumoto and Hikosaka, 2007, 2008). LHb neurons showed phasic activation in response to negative prediction errors and showed phasic inhibition when positive prediction errors occurred.

The habenula is located in the epithalamus, and is an evolutionarily conserved structure that exists in all vertebrates (Hikosaka, 2010; Stephenson-Jones et al., 2012). LHb has been shown to innervate neuromodulatory systems such as dopamine and serotonine neurons in the midbrain of both mammalian and non-mammalian species. In mammals, LHb sends both direct and indirect projections to dopamine, 5-HT and noradrenaline neurons in the midbrain (Araki et al., 1988; Herkenham and Nauta, 1979; Jhou et al., 2009a). This wide innervation of monoamine systems is reflected in the wide spectrum of behavior that the habenula affects. Since the 1950s, many lesion studies of the habenula complex (including medial habenula and LHb) have suggested its role in sleep (Haun et al., 1992; Valjakka et al., 1998), maternal behavior (Matthews-Felton et al., 1995), stress (Wirtshafter et al., 1994), and cognitive functions such as attention and spatial reference memory (Lecourtier and Kelly, 2005). In addition, abnormal increase of LHb activities has been indicated to cause symptoms similar to depression in mice (Li et al., 2011, 2013). Despite these extensive behavioral studies, it has been difficult to conclude what habenula is essential for. By recording the activities of lateral habenula neurons in
behaving monkeys, Hikosaka’s group discovered that LHb encodes reward prediction error. This finding suggested that it might play an important role in reinforcement learning.

Is the prediction error signal encoded by LHb relevant for behavior? Electrical stimulation of LHb while monkeys were performing a visually guided saccade task decreased the value of the visual target, as shown by a longer saccade latency of the animal in subsequent trials (Matsumoto and Hikosaka, 2011). Moreover, the increase of saccade latency occurred gradually after repeated stimulation, suggesting that the effect is due to learning rather than online modulation. Anatomically, LHb projects to VTA dopamine neurons through an inhibitory nucleus called the rostromedial tegmental area (RMTg) (Balcita-Pedicino et al., 2011; Jhou et al., 2009a, 2009b; Omelchenko et al., 2009), as well as GABA neurons in VTA (Omelchenko et al., 2009). Optogenetic activation of the projection from LHb to RMTg can cause conditioned place aversion (Stamatakis and Stuber, 2012). In an operant conditioning paradigm, mice increase nose poking frequency when the action produces a pause of LHb stimulation; conversely, mice decrease poking frequency when nose poking activates the LHb to RMTg pathway (Stamatakis and Stuber, 2012). Collectively, these results suggest that activating lateral habenula is sufficient to negatively reinforce behaviors.

In addition to the fact that LHb encodes negative prediction errors, several other lines of evidence also support the notion that dopamine neurons might inherit their ability to signal prediction errors from LHb. First, anatomically, there is a bisynaptic inhibitory projection from LHb to VTA dopamine neurons. LHb sends a glutamatergic projection to RMTg, and RMTg sends inhibitory projection to VTA dopamine neurons (Balcita-Pedicino et al., 2011; Jhou et al., 2009a, 2009b; Omelchenko et al., 2009). Secondly, electrical stimulation of LHb exerts potently inhibits onto dopamine neurons in vivo (Christoph et al., 1986; Ji and Shepard, 2007; Matsumoto
and Hikosaka, 2007). Lastly, the phasic excitation of LHb neurons to a cue predicting no reward precede dopamine neurons’ inhibition, although the inhibitory responses in lateral habenula to reward-predicting cues lagged behind the responses in dopamine neurons (Matsumoto and Hikosaka, 2007). These results suggest that dopamine neurons could get receive calculated prediction error signals from LHb rather than computing the signals themselves.

Although activation of LHb is sufficient to inhibit dopamine neurons and cause behavioral changes, it is not known whether habenula is necessary for normal RPE signaling in dopamine neurons. Given that dopamine neurons receive broad inputs from dozens of brain regions (Geisler and Zahm, 2005; Swanson, 2000; Watabe-Uchida et al., 2012), it is also possible that multiple mechanisms produce the computation of RPE signals in dopamine neurons. In Chapter 1, I test this hypothesis by lesioning the habenula and recording the activity of dopamine neurons in a classical conditioning task.

**Long range inputs of dopamine neurons**

In addition to receiving direct inputs from local VTA GABA neurons and the neighboring nucleus RMTg, dopamine neurons also receive long range inputs from many other brain areas. It has been difficult to trace direct inputs to dopamine neurons because traditional tracing methods label inputs to all VTA neurons including both dopaminergic and non-dopaminergic neurons. Recent developments in viral tracing tools have facilitated experiments that trace the monosynaptic inputs specifically to dopamine neurons.

By taking advantage of rabies virus tracing tools, Watabe-Uchida et al. found that dopamine neurons in the VTA receive inputs from areas. Many of these input areas are located along the medial forebrain bundle (Watabe-Uchida et al., 2012), including dorsal striatum (DS),
ventral striatum (VS), ventral pallidum (VP), lateral hypothalamus and subthalamic areas. The specificity of rabies virus labeling is achieved by limiting the expression of TVA and RG proteins only in dopamine neurons. TVA is a cognate receptor for pseudotyped rabies virus, and RG (rabies glycoprotein) is a critical membrane molecule for trans-synaptic spread of the rabies virus. In DAT-Cre mice, Cre is expressed under the control of the dopamine transporter (DAT) gene, therefore restricted to dopaminergic neurons (Bäckman et al., 2006; Lammel et al., 2015). By injecting two AAV helper viruses carrying TVA and RG respectively into DAT-Cre mice, the expression of TVA and RG is therefore restricted to dopamine neurons. Initial infection of rabies virus is restricted to dopamine neurons through TVA, and rabies virus can spread only to neurons that provide monosynaptic inputs to dopamine neurons through RG, since TVA and RG are only expressed in dopamine neurons (Wickersham et al., 2007a, 2007b), see **Figure 0.2**.

![Diagram of the modified rabies virus system](https://via.placeholder.com/150)

**Figure 0.2 A modified rabies virus system labels the monosynaptic inputs of dopamine neurons**

By using DAT-Cre mice and AAV carrying cre-inducible TVA and RG proteins, these proteins are only expressed in dopamine neurons (blue). As a result, rabies virus only infects dopamine
Figure 0.3 (Continued). neurons but not other local neurons (purple). The monosynaptic inputs of dopamine neurons are also labeled by rabies virus (cyan).

As previously discussed, the striatum is not only the main projection target of dopamine neurons, but also provides the largest number of inputs to dopamine neurons (Watabe-Uchida et al., 2012). Many theoretical models have hypothesized the striatum as the source of value signals in TD learning (see Equation 5) for several reasons: 1) it has been shown that neurons in the striatum encode reward expectation signals by showing sustained change of activities after the onset of reward predictive cues (Schultz et al., 1992; Tremblay et al., 1998), similar to the form of $V(t)$ signals in TD model, 2) In addition to directly projecting to dopamine neurons, the striatum is also indirectly connected to dopamine neurons through the ventral pallidum. This additional synapse could introduce a fixed delay and function as $V(t+1)$ upon combining the signals from striatum. By taking the difference of the striatum $V(t)$ signal and the pallidum $V(t+1)$ signal, and adding an excitatory input of $r(t)$, dopamine neurons can compute TD error signals, see Figure 0.1. It remains to be demonstrated whether dopamine neurons do receive $V(t)$ and $V(t+1)$ from these areas.

The lateral hypothalamus has long been implicated in reward processing. In a classic study of mapping reward centers, Olds and Milner identified the lateral hypothalamus as one of the hotspots that induced self-stimulation (Olds and Milner, 1954). Moreover, normal function of lateral hypothalamus is thought to be required for reward-seeking behavior. Both electrolytic and chemical lesions of lateral hypothalamus suppress feeding and drinking (Stuber and Wise, 2016). Lateral hypothalamus has also been shown to respond to reward predictive cues and delivery of reward (Nakamura and Ono, 1986; Nieh et al., 2015). Stimulation axon terminals of lateral
hypothalamus neurons synapsing in VTA causes excitations in ~55% of dopamine neurons and causes inhibitions in ~30% of dopamine neurons (Nieh et al., 2015). These responses are mediated by both a direct projection from LH to dopamine neurons as well as an indirect connection via VTA GABA neurons. Given the relatively strong excitatory connection from lateral hypothalamus neurons to dopamine neurons, it is possible that the lateral hypothalamus contributes to dopamine RPE by providing information about reward CS and actual reward.

PPTg is a nucleus located in the brain stem, posterior to VTA, and has been hypothesized to cause phasic burst firing in dopamine neurons (Grace et al., 2007). Stimulation of PPTg causes excitatory postsynaptic potentials in SNc dopamine neurons, and this effect depends on both glutamatergic and cholinergic inputs (Futami et al., 1995). Furthermore, stimulation of PPTg can cause bursts firing in putative dopamine neurons, resembling phasic responses to reward predictive cues and rewards in dopamine neurons (Floresco et al., 2003; Lokwan et al., 1999). Could PPTg provide the signals to drive dopamine neurons’ phasic excitation during normal behaviors? It has been shown that PPTg neurons respond to sensory predictive cues with different modalities in rats (Pan and Hyland, 2005). PPTg neurons has also been shown to respond to reward predictive cues and rewards in primates (Hong and Hikosaka, 2014; Kobayashi and Okada, 2007; Okada et al., 2009a). It remains to be tested whether those PPTg neurons encoding rewards or reward predictive cues are actually connected to dopamine neurons and whether these neurons contribute to RPE signals in dopamine neurons.

To survey how different input areas contribute to the RPE signals in dopamine neurons, in Chapter 2, we systematically mapped the information transmitted to dopamine neurons in each of the seven input areas (Watabe-Uchida et al., 2012) at single unit resolution. To record from neurons that directly synapse onto dopamine neurons, we combined the aforementioned rabies-
virus tracing approach with optogenetic identification techniques in \textit{in vivo} electrophysiological recordings. Furthermore, we used a well-controlled behavioral paradigm designed to probe the prediction error signals in awake recording conditions. This experimental design allows us for the first time to ask what information is conveyed to dopamine neurons by individual input neurons. In addition, comparing different input areas’ responses in the same behavioral task will provide in valuable information about how information about reward is processed in a distributed network across multiple brain areas.
CHAPTER 1

Habenula Lesions Reveal that Multiple Mechanisms Underlie Dopamine Prediction Errors

This chapter is based on a paper I published together Professor Uchida, entitled ‘Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors.’ I collected data and performed data analysis. I wrote the manuscript with comments from Professor Uchida.
Introduction

The ability to predict future outcomes based on sensory inputs is critical for making proper decisions. Psychological studies of animal learning have shown that temporal contiguity between two events (e.g. a sensory cue and reward) is not sufficient for establishing an association between them. Instead, it has been postulated that the efficiency of learning depends on how surprising the outcome is (Kamin, 1969; Rescorla and Wagner, 1972; Steinberg et al., 2013). According to Rescorla and Wagner (1972), learning ($\Delta v$) is proportional to the discrepancy between the value of obtained reward ($R$) and the predicted value of reward ($v$), or reward prediction error (RPE) ($R - v$):

$$\Delta v = \alpha(R - v)$$

Dopamine neurons in the midbrain signal RPEs (Bayer and Glimcher, 2005; Cohen et al., 2012; Schultz et al., 1997). Dopamine neurons are excited by unpredicted reward. When a sensory cue predicts a reward, dopamine neurons respond to the reward-predictive cue and their response to the predicted reward is greatly reduced. Furthermore, when a predicted reward is omitted, dopamine neurons decrease their firing (‘dip’) from baseline. The mechanisms by which dopamine neurons generate RPE signals, however, remain largely unknown.

Recent studies have found that neurons in the lateral habenula (LHb) also encode RPEs, but in the opposite direction compared to dopamine neurons (Matsumoto and Hikosaka, 2007, 2008), and that stimulation of LHb causes transient inhibition of dopamine neurons (Christoph et al., 1986; Ji and Shepard, 2007; Matsumoto and Hikosaka, 2007). Furthermore, stimulation of LHb neurons is sufficient to cause aversive learning (Matsumoto and Hikosaka, 2011; Stamatakis and Stuber, 2012). These findings raised the possibility that RPE signals are already calculated in LHb and simply relayed to dopamine neurons. Although intriguing, this hypothesis
has not been tested experimentally. Moreover, at the behavioral level, it is unclear whether and how the habenula regulates learning that depends on positive or negative RPEs.

To address these questions, we performed bilateral lesions of the habenula and examined the effects on the firing patterns of dopamine neurons as well as behavioral performance in a classical conditioning task. We found that habenula lesions significantly impaired dopamine neurons’ normal inhibitory response to reward omission. By contrast, inhibitory responses to air puff-predictive cues or air puffs remained intact in lesioned animals. Furthermore, dopamine neurons encoded RPE signals during reward-predictive cues and reward in lesioned animals, although these responses were less reliable. At the behavioral level, we observed a phenotype that is consistent with a relative reduction of negative, over positive, RPEs. Taken together, our results support the idea that multiple inputs play a role in the generation of RPE-related activities of dopamine neurons.

**Results**

**Behavioral paradigm and habenula lesions**

We trained mice in a classical conditioning paradigm in which odor cues (conditioned stimuli, CS) predicted either appetitive or aversive outcomes (unconditioned stimuli, US) with different probabilities (*Figures 1.1 A and B*). In reward trials, the size of water was constant and the probability of water delivery was varied (90, 50, 10%) for different odor cues. In aversive trials, air puffs were delivered to the face of the mouse with 90% probability. Each behavioral trial began with an odor cue (CS; 1s) followed by a 1s delay and an outcome (US). In addition to the cued trials, we interleaved reward alone (free reward) or air puff alone (free air puff) in some trials without an odor cue. Mice began to lick during the delay between the reward-predictive cue
and reward. Lick frequencies were higher in trials with higher reward probabilities, indicating that the animals learned the association of odor cues and the expected value of the upcoming reward (anticipatory licking for 100% reward > 50% reward > 0% reward; $P < 0.001$ for both, Wilcoxon signed rank test, $n = 12$ mice; Figure 1.C).

Figure 1.1 Odor-outcome association task
(A) Experimental set-up. (B) Task design. (C) Frequency of licking during pre-operative period (mean ± s.e.m., bins of two days).

To examine the role of habenula in the activity of dopamine neurons and behavior, we lesioned the habenular complex bilaterally in a set of animals ($n = 5$ mice), after initial conditioning training. To make the lesions of the habenula as complete as possible, we chose to perform electrolytic lesions (See Discussion and Methods). For comparison, another set of animals ($n = 7$ mice) underwent sham-lesions or no operations. Lesions covered a large portion of the habenula with occasional, small lesions in the medial part of the hippocampus and the paraventricular thalamic nucleus (PVT) (Figure 1.2, Table 1.1).
Figure 1.2 Electrolytic lesions in habenula and recording sites in VTA region

(A) Locations of electrolytic lesions. The lesion areas of each of the five habenula-lesioned animals are marked in red. Darker red color indicates that more animals have lesions in that area.

27
**Figure 1.2 (Continued).** (B) Recording sites in VTA. Each circle is for one animal. Black circles, control animals; red circles, lesion animals. Recordings were made in similar regions between control and lesion animals. (C) Examples of Nissl-stained slices. Left, a brain slice from one control animal; middle and right, example histology from four lesion animals, including both small and large lesions. Red lines indicate the boundary of lesioned area (see Methods). LHB, lateral habenula; MHB, medial habenula; D3V, dorsal third ventricle; PV, paraventricular thalamic nucleus; sm, stria medullaris.

| Table 1.1 Quantification of brain areas covered by lesion |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Fraction of lesion area**     | **Lateral habenula ipsilateral** | **Lateral habenula contralateral** | **Medial habenula** | **Hippocampus** | **Paraventricular thalamic nucleus** |
| mean ± s.d.                     | 86 ± 12 %       | 51 ± 24 %       | 96 ± 3 %        | 3 ± 1.5 %       | 30 ± 17 %       |
| min-max                         | 73 – 100 %      | 17 – 74 %       | 91 – 100 %      | 1 – 6 %         | 4 – 50 %        |

**Elevated and less discriminable anticipatory licking in lesion animals**

During the training period, there was no difference between lesion and control groups in anticipatory licking ($P = 0.095$, 3-way analysis of variance [ANOVA]). During the post-operative period, control animals’ lick rates were stable across days (**Figure 1.3 A**). However, lesion animals’ anticipatory licking in both 50% and 90% reward trials gradually increased over 15 days ($r = 0.89$, $P < 0.001$, 50% trials; $r = 0.77$, $P < 0.001$, 90% trials; Pearson correlation; **Figure 1.3 A**), resulting in significantly higher lick rates than control mice ($P < 0.0001$, 3-way ANOVA). Furthermore, the distribution of anticipatory licks in 50% reward trials became more similar to 90% reward trials in the lesion group (**Figure 1.3 B and C**). These results show that
the lick frequency became less sensitive to fine differences in the probability of upcoming reward.

![Graph A](image1.png)

**Figure 1.3 Habenula lesion altered the anticipatory licking behavior**

(A) Frequency of licking during post-operative period (mean ± s.e.m., bins of three days). (B) Histogram of anticipatory lick frequency. n, number of trials. (C) Ratio between anticipatory lick frequencies in trials with 50% and 90% probabilities of reward for early and late days after operation. *, P < 0.05; n.s, P > 0.05 (Wilcoxon rank-sum test). One control animal did not perform enough sessions during the 11-15 day period, thus was omitted from the analysis.

**Overall firing patterns of dopamine neurons in control and lesion animals**

We recorded the spiking activity of neurons in the VTA (170 neurons in 7 control animals and 276 neurons in 5 lesion animals, Figure 1.2 B; control: 14 ± 7 days, lesion: 15 ± 8 days, mean ± s.d.). To identify dopamine neurons while recording, we tagged dopamine neurons
with a light-gated cation channel, channelrhodopsin-2 (ChR2) (Methods). Dopamine neurons were identified based on their responses to light delivered through an optical fiber placed near the tip of the electrodes. In addition to reliable spiking responses to light, we also verified that the shape of light-evoked spikes was almost identical to that of spontaneous spikes (correlation coefficient > 0.9). Based on these criteria (see Methods for detail), we obtained 45 and 44 dopamine neurons in control and lesion animals, respectively (control: 6.4 ± 5.3, lesion: 8.8 ± 4.9 neurons per animal; mean ± s.d. Figures 1.4). These neurons showed short-latency spikes in response to light (3.4 ± 1.1 ms, mean ± s.d.), and had little jitter in the latency of the first spike (Figure 1.4 F), indicating that they were directly activated by light.

Figure 1.4 Optogenetic identification of dopamine neurons
(A) Voltage trace an example dopamine neuron. Cyan bars: laser stimulation. One spontaneous (left) and one light-triggered (right) spikes are shown below. (B) Raster plot of the same neuron to 10 Hz (left) and 50 Hz (right) stimulations. (C) Isolation of this neuron (arrow) from the noise cluster. The energy of spike waveform is defined as the integral of the squared voltage values (\(\int v^2 dt\)). (D) Histogram of p-values testing whether light-activation induced significant changes in
Figure 1.4 (Continued). spike timing (n = 446 units). The P-values were derived from SALT (Stimulus-Associated spike Latency Test; see Methods) (Kvitsiani et al., 2013). Neurons with P-values < 0.05 and waveform correlations > 0.9 were considered identified (grey). (E) Probability of a spike as a function of stimulation frequency for each dopamine neuron (grey) and the mean across dopamine neurons (blue). (F) Histogram of mean (left) and S.D. (right) spike latency to light stimulation.

Identified dopamine neurons in the control group showed characteristic firing patterns consistent with RPE signals (Figure 1.5 A, left). First, they were activated by reward-predictive CSs in a value-dependent manner. Second, they responded strongly to unexpected reward, and their reward responses were reduced when an odor cue predicted the reward. Third, when an odor predicted reward, omission of that reward caused a transient decrease (“dip”) in activity below baseline.

In lesion animals, the baseline firing rate of dopamine neurons was elevated (control: 5.60 ± 2.26 spikes/s; lesion: 6.64 ± 3.20 spikes/s; mean ± s.d.; P < 0.05; Wilcoxon rank sum test; Figure 1.5 B), but did not correlate with days of recording (Figure 1.5 C). Dopamine neurons largely maintained their phasic response patterns (Figure 1.5 A, right). However, the fraction of neurons responding to particular events and the magnitude of these responses were altered. We will quantify these results in greater detail in the following sections.
Figure 1.5 The overall firing patterns of dopamine neurons

(A) Average firing rates of all identified dopamine neurons recorded in control (left) and lesioned (right) animals. US+, Trials in which outcome was delivered; US-, Trials in which outcome was omitted. Grey area indicates the time of odor stimulation. Dash line indicates expected US onset time. n, number of neurons from 7 control and 5 lesion animals. The same sets of neurons are used for (B) and (C). (B) Boxplot of baseline firing rate of dopamine neurons. The edges of the boxes are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers. Points are drawn as outliers if they are larger than Q3+1.5*(Q3-Q1) or smaller than Q1-1.5*(Q3-Q1), where Q1 and Q3 are the 25th and 75th percentiles. Outliers were plotted as individual data points. *, P < 0.05 (Wilcoxon rank-sum test).

(C) Baseline firing rate of individual dopamine neurons plotted as a function of days of recording. The firing rate of neurons in lesion group is not significantly correlated with recording days (r = 0.24, P = 0.11; Pearson correlation). Black, control; red, lesion.
Inhibitory responses during reward omission were greatly reduced in lesion animals

In control animals, reward omission caused a significant dip in most dopamine neurons (see Figure 1.5 A for average responses and Figure 1.6 A for example neuron). The response of individual dopamine neurons was visualized using the receiver-operating characteristic (ROC) analysis (Green and Swets, 1966). For each neuron, firing rate change from baseline was quantified using the area under the ROC curve (auROC) in a sliding window. Values greater than 0.5 indicate increases in firing rate and values less than 0.5 indicate decreases (shown by yellow and blue, respectively, in Figure 1.6 B). In control animals, many individual dopamine neurons showed a transient decrease in firing rate around the time when a reward was expected. In lesion animals, dopamine neurons’ dip during reward omission became sporadic in timing and less prominent over the population (Figure 1.6 B).

In control animals, 86.7% of the dopamine neurons showed a significant dip in activity during omission of 90% reward ($P < 0.05$, Wilcoxon rank sum test), while only 47.7% in lesion animals did so (Figure 1.6 C). The magnitude of the dip, as quantified by the decrease in firing rate from baseline during omission of 90% reward, was also reduced in lesion animals (control: $-2.6 \pm 0.2$, spikes/s; lesion: $-1.5 \pm 0.2$, spikes/s; mean $\pm$ s.e.m; $P < 0.0001$, Wilcoxon rank sum test; Figure 1.6 D). Furthermore, we found that the firing rate difference between the omission of 50% reward versus that of 90% reward was smaller in lesion animals (firing rate difference: control: $P < 0.001$, lesion: $P > 0.05$, Wilcoxon signed rank test; Figure 1.6 D; auROC, 90% versus 50% reward omission: control, $0.30 \pm 0.13$; lesion, $0.46 \pm 0.18$; mean $\pm$ s.d; $P < 0.0001$, Wilcoxon rank sum test; Figure 1.6 E; see Methods). We also noticed that dopamine neurons began decreasing their firing rate before the expected timing of reward delivery. This is probably due to uncertainty in reward timing because the timing at which an odor reaches the olfactory
epithelium may vary depending on the timing of inhalation onset. We found that this “pre-reward dip” was affected by lesions in a similar manner as the dip after the time of expected reward onset (Figure 1.7).

We examined whether unintended lesion of PVT contributed to the lesion effect; we did not find significant correlations of the PVT lesion size and reward omission response ($r = -0.19$, $P = 0.21$; Pearson correlation). Together, these results demonstrate that the inhibitory responses in dopamine neurons during reward omission were impaired by lesioning the habenular complex.

**Figure 1.6** Inhibitory responses during reward omission were diminished in lesion animals

34
Figure 1.6 (Continued). (A) Example neuron’s responses during reward omission. (B) Temporal profiles of all dopamine neurons during omission of 90% reward. Colors indicate an increase (yellow) or decrease (blue). (C) Percentage of neurons that showed a significant response to reward omission (versus baseline, P < 0.05, Wilcoxon signed-rank test). ***, P < 0.001; *, P < 0.05 (Binomial test). Analysis window is 2-3s after odor onset, same for D and E. Black, control; red, lesion. (D) Response magnitude during reward omission as measured by firing rate changes from baseline (mean ± s.e.m.). Blue asterisks indicate a significant difference between control and lesion groups (***, P < 0.001; no labeling, P > 0.05; Wilcoxon rank-sum test). Black and red asterisks indicate a significant difference between responses to two probabilities of reward within the control or lesion group (***, P < 0.001; n.s., not significant; Wilcoxon signed-rank test). (E) Boxplot of the auROC values (omission of 90% reward versus omission of 50% reward). ***, P < 0.001 (Wilcoxon rank-sum test).

Figure 1.7 Dopamine neurons’ response in pre-reward period
(A) Percentage of neurons that showed a significant response to pre-reward dip (versus baseline, P < 0.05, Wilcoxon signed-rank test). Asterisks indicate a significant difference between control and lesion groups (***, P < 0.001; Binomial test). Analysis window is 1.8-2s after odor onset,
**Figure 1.7 (Continued).** same for B and C. Black, control; red, lesion. (B) Response magnitude during pre-reward dip as measured by firing rate changes from baseline (mean ± s.e.m.). Blue asterisks indicate a significant difference between control and lesion groups (***, P < 0.001; no labeling, P > 0.05; Wilcoxon rank-sum test). Black and red asterisks indicate a significant difference between responses to two probabilities of reward within the control or lesion group (***, P < 0.001; n.s., not significant; Wilcoxon signed-rank test). (C) Boxplot of the auROC values (90% reward trials versus 50% reward trials). Values in the lesion group are significantly greater than control group (***, P < 0.001; Wilcoxon rank-sum test).

**Inhibitory responses to air puff remain unimpaired**

LHb neurons are excited by aversive outcomes and cues that predict them, in addition to the omission of predicted reward (Matsumoto and Hikosaka, 2008). If habenula plays a general role in generating the inhibitory responses of dopamine neurons, dopamine neurons’ inhibition by air puff-predictive cues and air puff should also be reduced.

Many identified dopamine neurons showed biphasic responses to air puff, consisting typically of a brief excitation followed by inhibition (**Figures 1.8 A, B**). In both excitatory and inhibitory phases, an unpredicted (“free”) air puff caused significantly stronger responses than did a predicted air puff (**Figure 1.9**). We first quantified the net response of individual neurons using 0-400 ms time window to cover the entire air puff response period. Contrary to the above prediction, a larger fraction of neurons in the lesion group were significantly inhibited by a free air puff than in the control group (control: 18%, lesion: 36%, P <0.05, Wilcoxon signed rank test; the difference in the fractions between control and lesion animals: P < 0.05, Binomial test; **Figure 1.8 C**). Similar trends were observed when the responses were quantified using firing rate changes (**Figure 1.8 D**). We also examined the prediction error coding for air puff by
quantifying the discriminability between predicted and unpredicted air puff using the auROC values; there was no significant difference between the control and lesion group (Figure 1.8 E).

**Figure 1.8 Inhibitory responses to air puff-predictive cues and air puff were unimpaired.**

(A) Average firing rates in air puff trials. Grey area, CS period. Black bars, time windows for data analysis in (C)-(E). n, number of neurons; n = 37 for control group’s response to free airpuff; same set of neurons is plotted for (C)-(F). (B) Temporal profiles of individual dopamine neurons in air puff trials using a sliding-window auROC analysis (time bin: 100ms; against baseline). The same neuron’s response to free air puff (unpredicted air puff) is shown on the right. White areas on the right are sessions without free air puff trials. (C) Percentage of neurons that showed
Figure 1.8 (Continued). a significant response to air puff or air puff-predictive cues. Black, control; red, lesion. Empty bar, significant excitation (versus pre-odor baseline, P < 0.05, Wilcoxon signed-rank test); filled bar, significant inhibition (versus pre-odor baseline, P < 0.05, Wilcoxon signed-rank test). *, P < 0.05 (Binomial test). (D) Comparison of air puff response amplitudes between control and lesion group, measured by firing rate changes from baseline. *, P < 0.05 (Wilcoxon rank-sum test). (E) Boxplot of auROC (90% airpuff versus free air puf).
Figure 1.9 (Continued). as most dots are above the diagonal (y=x). In lesion group, one out-of-range data point (x = 7.8, y = 31.2) is not shown. P-values are calculated using Wilcoxon signed-rank test. Black, control, n = 37 neurons; red, lesion, n = 44 neurons. (B) Response during the later inhibition phase (200-400 ms). Responses to free air puff are more inhibited than 90%-predicted air puff, as most dots are below the diagonal (y=x). We used a time window of 200-400ms because this corresponds to a period of purer inhibitory response (during 120-200ms, there are still many neurons excited by air puff). The sets of neurons and statistics are the same as those in (A).

Figure 1.10 Response to air puff in different time windows
**Figure 1.10 (Continued).** (A) Air puff response in the early excitation time window (0-120 ms). In this early time window, dopamine neurons in the control and lesion group are not significantly different in terms of the percentage of neurons showing significant response (P > 0.05, binomial test; left) or amplitude of response (P > 0.05, Wilcoxon rank sum test; right). Black, control, n = 45 neurons (37 neurons for free air puff); red, lesion, n = 44 neurons. A significance of response was judged using a Wilcoxon signed-rank test (versus baseline; P < 0.05). (B) Air puff response in later inhibition time window (120-400 ms), plotted in the same way as (a). dopamine neurons in the lesion group were more inhibited by air puff, both in terms of percentage of neurons showing significant inhibition (P < 0.001, Binomial test; left), or in the amplitude of response (P < 0.01, Wilcoxon rank-sum test; right). The same set of neurons and statistical tests are used as in (A).

We next analyzed the data by dividing the analysis window into early excitatory and late inhibitory epochs based on the average time course in control animals (0-120 ms and 120-400 ms from air puff onset for excitatory and inhibitory periods, respectively). During the early excitation epoch, we observed no significant difference in neurons’ responses to predicted or free air puffs (**Figure 1.10 A**). In the late inhibitory response epoch, neurons in lesion animals were significantly more inhibited than those in control animals (**Figure 1.10 B**). These results hold true even if we shifted the boundary between early and late analysis windows (e.g. 100 ms or 150 ms).

During the CS period, about half of identified dopamine neurons showed a decrease in firing in response to air puff-predictive cues compared with pre-odor baseline activity (**Figure 1.8 C, D**). These inhibitory responses were also unimpaired by habenula lesions. In summary,
these results show that habenula lesions preferentially impaired the response to reward omission while leaving other inhibitory responses either unaffected or even increased.

**Positive RPEs were weakened but preserved**

We next analyzed phasic excitation following positive RPE events including reward-predictive cues and reward (Figure 1.11). In control animals, almost all identified dopamine neurons showed a transient excitation in response to the 90% reward CS (95.6%, 43 of 45 dopamine neurons; spike counts in a 0-600 ms window after CS onset against baseline; $P < 0.05$, Wilcoxon rank sum test; Figure 1.11 A) or to unexpected reward (100%, 45 of 45 dopamine neurons; spike counts in a 0-400 ms window after US onset against baseline, $P < 0.05$, Wilcoxon rank sum test; Figure 1.11 B). By contrast, in lesion animals, fewer neurons responded significantly to 90% reward CS (70.5%, 31 of 44 dopamine neurons; Figure 1.11 A) or to unexpected reward (84.1%, 37 of 44 dopamine neurons; Figure 1.11 B). When the magnitude of the responses was compared, phasic firing to events with positive RPE signals was smaller in the lesion group (54% decrease for 90% reward predictive cue; 38% decrease for free or 10% reward; Figure 1.11 C, D). These CS and US responses were also less reliable in distinguishing 90% versus. 50% probability of reward in lesion animals (auROC, 90% versus 50% reward, $P < 0.0001$ for both CS and US responses when comparing control and lesion group, Wilcoxon rank sum test; Figure 1.11 E, F).

Although dopamine neurons in the lesion group signaled positive RPE less reliably, these responses were still modulated by reward expectation ($P < 0.001$, Wilcoxon signed rank test; Figure 1.11 C, D), exhibiting the hallmark of RPE-related activity. This is true even in the animals with almost complete lesions of the habenula (Figure 1.12). These results suggest that
the habenula boosts the positive RPE-related responses of dopamine neurons, but may not be required for these responses.

**Figure 1.11 Phasic excitations to reward CS and US were weakened in lesion animals**

(A) Percent of neurons that showed a significant response (versus baseline, P < 0.05, Wilcoxon signed-rank test) during reward CS (0-600ms after CS onset). ***, P < 0.001; *, P < 0.05; n.s, not significant (Binomial test). Black, control (n = 45 neurons, 7 mice); red, lesion (n = 44 neurons, 5 mice). The same sets of neurons are plotted for (B)-(F). (B) Percent of neurons that showed a significant response during reward (0-400ms after US onset). Free reward trials were not significantly different from 10%-predicted-reward trials, and these trials types were combined. (C, D) The magnitude of response (mean ± s.e.m.) to reward CS (C) and reward US (D), subtracted by baseline before trial starts. Filled symbols indicate a significant deviation from zero (P < 0.05, Wilcoxon signed-rank test). Blue asterisks indicate a significant difference
**Figure 1.11 (Continued).** Between control group and lesion group (***, P < 0.001; **, P < 0.01; no labeling, P > 0.05, Wilcoxon rank-sum test). Black and red asterisks indicate a significant difference within control or lesion group to different probabilities of reward (***, P < 0.001; **, P < 0.01; *, P < 0.05; Wilcoxon signed-rank test). (E, F) Boxplot of the auROC values (90% versus 50% reward trials) during reward CS (E) and reward US (F). Asterisks indicate a significant difference between control group and lesion group (***, P < 0.001; n.s, P > 0.05, Wilcoxon rank-sum test).

**Figure 1.12** Dopamine neurons’ response from animals with large lesions.
Figure 1.12 (Continued). (A) Average response patterns of dopamine neurons from two animals with large lesions. Note the phasic excitations in dopamine neurons are still prominent. n, number of neurons, from two animals. (B, C) The magnitude of response (mean ± s.e.m.) to reward CS (B) and reward US (C), subtracted by baseline before trial start. Filled symbols indicate a significant deviation from zero (P < 0.05, Wilcoxon signed-rank test). Black and red asterisks indicate a significant difference within control or lesion group for different probabilities of reward (***, P < 0.001; Wilcoxon signed-rank test).

Analysis of putative GABA and other unidentified VTA neurons

VTA contains a large number of GABA neurons, which are directly innervated by excitatory projections from the LHb (Brinschwitz et al., 2010; Omelchenko et al., 2009). The firing patterns of VTA neurons in a similar task can be classified into three distinct clusters using an unsupervised method (Cohen et al., 2012). Dopamine neurons and GABA neurons corresponded to two of the three types identified with this method. Although we did not directly identify GABA neurons in the present study, we analyzed the putative GABA neurons by classifying neurons into three response types. First, we clustered all recorded VTA neurons in control and lesion animals into three clusters based on their firing patterns in 90%-reward trials (Figure 1.13 A). All identified dopamine neurons in control animals and most (37/44) identified dopamine neurons in lesion animals fell into the first cluster, consistent with our previous study (Cohen et al., 2012) (Figure 1.13 A, B). Note that cluster 1 neurons in control group included a small fraction of neurons that were activated by air puff more strongly than any optogenetically-identified dopamine neurons; this population, which is likely to be non-dopaminergic, was largely absent in the lesion group (see Methods for details). Neurons in the second cluster showed sustained excitation after CS onset whose magnitude monotonically increased with
reward probabilities, similar to optogenetically-identified GABA neurons (Cohen et al., 2012). In addition, there was a third cluster of neurons that were inhibited during the delay.

Figure 1.13 Comparison of all VTA neurons’ responses in control and lesion group
(A) Clustering of response profiles. All recorded neurons are clustered into three groups based on their response profiles in 90% reward trials. The same neurons’ responses in 90% air puff trials are shown in the right panel. Clusters are separated by red lines. From top to bottom, cluster 1, 2 and 3 in (B) respectively. (B) Average firing patterns of three neuronal clusters in (A). Colors are as in Figure 1.5 A. Only US+ trials are shown. n, number of neurons. The same sets of neurons are plotted in (C)-(F). (C) Baseline firing rates (mean ± s.e.m.). ***, P < 0.001
Figure 1.13 (Continued). (Wilcoxon rank-sum test). cl1, cluster 1; cl2, cluster 2; cl3, cluster 3. (D) Response to air puff (0-200ms after air puff onset) (mean ± s.e.m., baseline subtracted). ***, P < 0.001; **, P < 0.01; n.s., not significant (Wilcoxon rank-sum test). (E) Delay period activity (auROC against baseline, 1-2s after odor onset) in 90% reward trials. ***, P < 0.001; n.s., not significant (Wilcoxon rank-sum test). (F) Delay period activity (1-2s after odor onset) in 90% reward trials versus 50% reward trials as quantified using the auROC against each other. ***, P < 0.001 (Wilcoxon rank-sum test).

Consistent with the habenula’s disynaptic inhibitory connections to dopamine neurons and direct excitatory connections to VTA GABA neurons, habenula lesions caused a slight increase in the baseline firing rates of neurons in cluster 1 and a decrease in cluster 2 (Figure 1.13 C). In addition to changes in baseline, habenula lesions also altered the task-related responses of neurons in cluster 2 (putative GABA neurons). First, whereas air puff caused phasic excitation in cluster 2 neurons of control group, these responses were dramatically decreased in lesion group (Figure 1.13 B, D). In addition, the robustness of sustained excitation during the delay in 90%-reward trials was greatly reduced in lesion animals (P < 0.0001, Wilcoxon rank sum test; Figure 1.13 E). The sustained excitations in 90% and 50% reward trials were less discriminable (P = 0.0017, Wilcoxon rank sum test; Figure 1.13 F).

**Behavioral phenotype is consistent with a relative decrease of negative over positive RPE**

The above results demonstrate that habenula lesions impaired aspects of RPE-related responses of dopamine neurons. In control animals, lick frequency in 50%-reward-probability
trials reached an asymptote at an intermediate level. This can be parsimoniously explained by RPE: in 50%-reward-probability trials, animals receive positive RPEs when reward was delivered and negative RPEs when reward was omitted. As a result, the predicted value of reward reaches equilibrium. We hypothesized that the elevated lick frequency in 50%-reward trials in lesion animals can be explained by unbalanced RPEs.

To test this idea, we implemented a simple reinforcement learning model (Rescorla and Wagner, 1972). In this model, animals learn to predict the value of upcoming rewards associated with different odor cues. The value of each odor was updated based on the magnitude of RPE multiplied by a learning rate parameter (\( \alpha \)). To dissociate the effect of positive versus negative RPEs in learning, two learning rate parameters (\( \alpha_p \) and \( \alpha_N \)) were assigned separately for each (Figure 1.14 A). The third parameter, \( R \), corresponds to the value of the outcome.

We first fit the model to the data in control animals during training and during the post-operative period. We found that the fitted value for \( \alpha_p \) (0.0020 \( \pm \) 0.0002) was very close to the fitted value for \( \alpha_N \) (0.0022 \( \pm \) 0.0003). The fitted value of \( R \) (6.8 \( \pm \) 0.3 licks/s) reflected the asymptotic lick frequency in high reward probability trials. Next, assuming that the lesion group has the same value of \( R \), we examined what combinations of \( \alpha_p \) and \( \alpha_N \) best matched the behavioral data in the lesion group. Given this model, we found that the probability of obtaining the observed behavioral data in lesion group was highest when the ratio between \( \alpha_p \) and \( \alpha_N \) was 0.0055 to 0.0004, while similar value of \( \alpha_p \) and \( \alpha_N \) best predicted the control data (Figure 1.14 B, C). Even when \( R \) was allowed to be different in the control and lesion group, we obtained consistent results that optimal fitting was achieved only when \( \alpha_p > \alpha_N \). In summary, the behavioral phenotype is consistent with a relative reduction of negative RPE-based learning over positive RPE-based learning.
Figure 1.14 Simulation of anticipatory licking based on the Rescorla-Wagner model

(A) Schematic of the learning algorithm. $V_N$, associative strength between odor cue and reward at trial N, corresponds to anticipatory lick frequency measured in trial N; $\alpha_P$, learning rate from positive prediction error; $\alpha_N$, learning rate from negative prediction error. The value of $V_N$ is updated in a trial-by-trial manner depending on whether reward is delivered or not, as well as the current value of $V_N$. (B) The quality of model fitting, indicated by log likelihood of observing the
**Figure 1.14 (Continued).** experimental data, depending on the combination of $\alpha_P$ and $\alpha_N$. Control group has the best fit when $\alpha_P$ and $\alpha_N$ have similar values (left); lesion group has the best fit above the diagonal line (i.e. $\alpha_P > \alpha_N$). Best fit parameters ($\alpha_P$ and $\alpha_N$) are marked by asterisks and correspond to the conditions for data fitting in (C). Due to differences in the duration of experiments, 16 days of data is used for control data fitting and 27 days of data is used for lesion group. Using 16 days of the data to fit both groups yielded qualitatively similar results. Same values of $R$ (6.8 licks/s) are used for control and lesion group. (C) Example fitted data from (B) using best fitted parameters.

**Discussion**

In the present study, we examined the role of the habenular complex (including both medial and lateral habenula) in RPE signaling of VTA dopamine neurons and in behavioral performance. Our results demonstrate that various aspects of RPE were impaired in habenula lesion animals. We found that the inhibitory responses of dopamine neurons upon reward omission were greatly diminished although inhibitory responses to aversive stimuli were relatively unimpaired. On the other hand, the effects on excitatory responses were much milder. These results suggest that the habenula contributes in generating dopamine responses, but only to particular events; other inputs are likely responsible for responses to other events.

**Technical considerations**

The present study was designed to examine the necessity of the habenular complex for reward prediction error signals in dopamine neurons. The habenula in mice is a longitudinal structure extending about 1 mm along the anterior-posterior axis and is located immediately adjacent to the third ventricle. In order to reliably inactivate the whole structure, we chose
electrolytic lesions over other methods such as excitotoxic lesions, optogenetics, pharmacogenetics or local pharmacology. These alternative methods run the risk of partial inactivation. Further, given the high baseline firing rates of habenula neurons, transient inactivation might cause a great increase in the baseline firing rates of dopamine neurons, which would in turn make it difficult to isolate the effects on phasic responses. For our purpose, permanent lesions have certain advantages. First, inactivation is complete, and it is easier to confirm the extent of inactivation (lesions) by histology. This method can be applied to an elongated brain area, and it is expected to be less prone to causing large changes in baseline firing in dopamine neurons. On the other hand, lesions have disadvantages. For instance, long-term compensation may occur over time (discussed further below).

After lesions of the habenula, dopamine neurons were identified using an optogenetic tagging method. This type of unambiguous identification was necessary because lesions could have dramatically altered the firing patterns of dopamine neurons. Furthermore, spike waveforms, a commonly used method to identify dopamine neurons, have been shown to be unreliable in recording conditions similar to the present study (Cohen et al., 2012).

Although the habenula is relatively isolated from other surrounding areas, it is possible that our habenula lesions might have damaged fibers of passage in the stria medullaris and fasciculus retroflexus. The stria medullaris mainly consists of afferent fibers to the habenula from forebrain areas such as the lateral hypothalamus, nucleus of diagonal band, septum nucleus and endopedunculus nucleus (Klemm, 2004). The fasciculus retroflexus is a fiber tract linking the habenula and subcortical areas such as the VTA, rostromedial tegmental nucleus (RMTg), raphe nucleus and interpeduncular nucleus (Araki et al., 1988; Herkenham and Nauta, 1979). Although we cannot exclude the possibility that we lesioned a fraction of these fibers that just...
passed the habenula, most of these fibers constitute inputs and outputs of the habenula. In addition, in some cases, we also damaged the dorsal part of the PVT. However, the size of PVT lesions was not correlated with the deficit in dopamine “dip” during reward omission. Although we cannot exclude the possibility that these unintended lesions contributed to the effect of lesions, in the following, we will provide a parsimonious explanation for our data based on reported neuronal activities and projections of neurons in the habenular complex (Hong et al., 2011; Jhou et al., 2009a; Matsumoto and Hikosaka, 2007, 2009).

**Mechanisms that generate RPE-related activity of dopamine neurons**

It had been proposed that LHb neurons send a relatively complete set of RPE signals to dopamine neurons (Hong et al., 2011; Matsumoto and Hikosaka, 2007). Our results partially support this idea, as habenula lesions affected phasic inhibitory responses during reward omission as well as excitatory responses to reward-predictive cues and reward. However, our results also suggest that multiple inputs underlie the generation of RPE-related activities of dopamine neurons. First, inhibitory responses to air puff and air puff-predictive cues remained intact even after lesions. Although we cannot completely rule out the possibility that the lack of effect was due to a compensatory mechanism after lesions, our results are consistent with the idea that inhibitions caused by reward omission involve different mechanisms than inhibitions caused by air puff or air puff predictive cues. In other words, if both inhibitory responses during reward omission and air puff are caused by the same mechanism, compensation should occur in a same degree to both of these responses. This was not the case. Second, in contrast to the results for reward omission, the effects on positive RPE were less evident. Specifically, in the lesion group, the responses to reward predictive cues and rewards were still significantly modulated by
expectation, while the responses to reward omission failed to distinguish high (90%) and medium probability (50%). In addition, a higher percentage of dopamine neurons in habenula-lesioned animals still showed significant responses to positive RPE than to reward omission. Even in the animals with more than 85% lesions of habenula, the phasic excitation remained largely intact (Figure 1.12). These results suggest that excitation of dopamine neurons requires other inputs than those from the habenula. This is consistent with a previous study which showed that LHb neurons have longer latency than dopamine neurons in responding to reward-predictive cues or reward (Matsumoto and Hikosaka, 2007). Thus, the habenula appears responsible mainly for the inhibitory responses of dopamine neurons to reward omission, but not to aversive stimuli, and only partially to reward. Taken together, these results suggest that the RPE-related activities of dopamine neurons are generated by multiple inputs.

The habenula can influence firing patterns of dopamine neurons through multiple pathways. Neurons in the LHb send glutamatergic projections to the RMTg, which contains GABA neurons synapsing onto dopamine neurons (Hong et al., 2011; Jhou et al., 2009a). It should also be noted that there are other pathways by which habenula neurons influence dopamine neurons. First, LHb neurons send direct projections to VTA GABA neurons as well as, to a lesser extent, to dopamine neurons (Jhou et al., 2009a; Omelchenko et al., 2009). Interestingly, cluster 2 (putative VTA GABA) neurons drastically reduced its activity after habenula lesions. The decreased baseline firing of VTA GABA neurons may, in turn, increase the baseline activity of dopamine neurons. Furthermore, the habenula projects to the dorsal raphe nucleus (Ogawa et al., 2014; Pollak Dorocic et al., 2014; Sutherland, 1982; Weissbourd et al., 2014), which is a major source of monosynaptic inputs to VTA dopamine neurons (Watabe-Uchida et al., 2012). Lastly, a less studied pathway is the medial habenula’s projection to the
interpeduncular nucleus (IPN) (Viswanath et al., 2013). Lesions of this pathway lead to an increase in dopamine levels (Nishikawa et al., 1986), which might contribute partly to the increase of baseline firing. Recording from the medial and lateral habenula in similar tasks will elucidate its function in reward processing. Furthermore, projection-specific manipulation of activity (Felix-Ortiz et al., 2013; Stamatakis and Stuber, 2012; Tye et al., 2011) will be a powerful means with which to uncover the significance of each of these pathways from the habenula to dopamine neurons.

**Neural mechanisms underlying prediction error-driven learning**

Prediction error-based learning allows an animal to learn the proper values associated with different stimuli (Dayan and Abbott, 2001; Rescorla and Wagner, 1972). Optimal learning requires a correct balance between learning from positive and negative RPEs. After habenula lesions, licking in 50% and 90% reward trials was elevated, which could be explained by the effect of weakened negative RPEs compared to positive RPEs. This is consistent with the decrease in negative RPEs and the comparably mild change in positive RPEs in dopamine neurons. Alternatively, licking might be increased because of a change in motivation. For example, dopamine neurons’ phasic response to reward-predictive cues as well as tonic firing have been linked to motivation (Niv et al., 2007). However, these mechanisms cannot explain the behavioral change. First, cue-evoked phasic responses of dopamine neurons were reduced overall after lesions (Figure 1.11 C). Second, although licking gradually increased over several days, the baseline firing did not increase over this period (Figure 1.5 C).

The behavioral changes by lesion of habenula may, however, be mediated by other mechanisms than dopamine neurons. First, it remains less evident whether a brief dip in
dopamine firing during reward omission is sufficient to cause learning from negative RPEs because previous studies often used relatively long inactivation of dopamine neurons compared to the natural dip (Danjo et al., 2014; Tan et al., 2012). Second, the habenula also has strong projections to the median as well as the dorsal raphe nuclei (Aghajanian and Wang, 1977; Kalén et al., 1989; Lecourtier and Kelly, 2007; Ogawa et al., 2014), which has been implicated in learning from negative events (Agetsuma et al., 2010; Amo et al., 2014; Cohen et al., 2015). Thus, impaired negative RPE-based learning in behavior could arise from pathways independent of dopamine neurons. Further studies are needed to elucidate the pathways through which the habenula controls RPE-based learning in mammals.
Methods

Animals

All procedures were carried out in accordance with NIH standards and approved by Harvard University Institutional Animal Care and Use Committee (IACUC). We used 13 adult male mice, backcrossed with C57/BL6 mice, heterozygous for Cre recombinase under the control of the DAT gene (B6.SJL-Slc6a3tm1.1(cre)Bkmn/J, the Jackson Laboratory)(Bäckman et al., 2006). Five animals in the habenula lesion group were verified by histology. Seven animals were in the control group including two with sham-lesion operation, one with only small contra-lateral side lesion of the media habenula, and four animals without operations in the habenula. Animals were singly housed on a 12-h dark/12-h light cycle.

Surgery and viral injections

Mice were surgically implanted with a custom-made metal plate (a head plate). During the same surgery, 500 -1000 nl adeno-associated virus (AAV), serotype 5, carrying an inverted ChR2 (H134R)-EYFP flanked by double loxP sites (Atasoy et al., 2008; Cohen et al., 2012) was injected into the VTA (from bregma: 3.1 mm posterior, 0.7 mm lateral, 4–4.2 mm ventral). The expression of this virus in dopamine neurons is highly selective and efficient, and ChR2 expression is uniform across dopamine neurons with different projection targets (Cohen et al., 2012; Lammel et al., 2015).

After 10 days of training on the conditioning task, mice were randomly selected to be in lesion or sham-lesion group. Electrolytic lesions were made bilaterally using a stainless steel electrode (15kΩ, MicroProbes, MS301G). Each side of the brain was lesioned at two locations (from bregma: 1.6 mm/1.9 mm posterior, 1.15 mm lateral, 2.93 mm depth, with a 14 degree
angle). A cathodal current of 150 μA was applied for 75s at bregma -1.6mm and 90s at bregma -1.9 mm. The head plate attached to the skull was used as the anode. For sham-lesion operations, no current was applied. In the same surgery, after lesions were made, a microdrive containing electrodes and an optical fiber was implanted in the VTA (from bregma: 3.1 mm posterior, 0.7 mm lateral, 3.8-4.0 mm ventral).

All surgery was performed under aseptic conditions with animals either under ketamine/medetomidine anesthesia (60/0.5 mg kg$^{-1}$, intraperitoneal, respectively) or isoflurane inhalation anesthesia (1-2% at 0.5-1.0 L/min). Analgesics (ketoprofen, 1.3 mg kg$^{-1}$ intraperitoneal, and buprenorphine, 0.1mg kg$^{-1}$ intraperitoneal) were administered postoperatively.

**Behavioral task**

After >1 week of recovery, mice were water-deprived. The body weight was maintained above 85% of their full body weight. Animals were head-restrained using a head plate and habituated for ~15 min for 1–2 d before training on the task. Odors were delivered with a custom-made olfactometer (Uchida and Mainen, 2003). Odors were isoamyl acetate, eugenol, 1-hexanol, p-cymene, ethyl butyrate, 1-butanol, and carvone (1/10 dilution in paraffin oil). A set of odor was assigned randomly for each animal. Licks were detected by breaks of an infrared beam placed in front of the water tube. Behavioral signals are digitized and recorded at 1kHz (PCI-6251, National Instruments).

During the training period, each odor predicted a drop of water (3.75 µl; valve open for 70ms) with different probabilities: 100%, 50%, nothing, or air puff delivered to the animal's face. The strength of air puff was enough to cause blinking behavior and was shown to be aversive in
a previous study (Cohen et al., 2015). Air puff trials were added after conditioning only with water for about 3-4 days. To measure the responses when actual outcome violates the expectation, during the recording sessions, we changed the reward probabilities to 90%, 50% and 10% and air puff probability to 10%. Inter-trial intervals (ITIs) were drawn from an exponential distribution, resulting in a flat ITI hazard function. Data from control mice were obtained from 61 sessions (1–10 sessions per animal, 9 ± 7 sessions); data from lesion mice were obtained from 77 sessions (9-19 sessions per animal, 15 ± 5 sessions). Animals performed between 300 and 600 trials per day (441 ± 82 trials.).

**Electrophysiology**

Recordings were made using a custom-built 200-μm-fibreoptic-coupled screw-driven microdrive with eight implanted tetrodes. Tetrodes were glued to the fiber optic with epoxy. The ends of the tetrodes were 350–600 μm from the end of fiber optic. Neural signals were amplified 200-fold with a filter between 0.1 and 9,000 Hz (RHA2116, Intan Technologies LLC), and digitized at 30kHz (PCle-6351, National Instruments). To extract timing of spikes, signals were band-pass-filtered between 300 and 6,000 Hz. Spikes were sorted offline using MClust-3.5 software (David Redish). At the end of each session, the fiber and tetrodes were lowered by 40-80μm to record new neurons next day.

To identify dopamine neurons, we used ChR2 to observe stimulation-locked spikes (Cohen et al., 2012; Jennings and Stuber, 2014; Kvitsiani et al., 2013; Lima et al., 2009). The optical fiber was coupled with a diode-pumped solid-state laser with analogue amplitude modulation (Laserglow Technologies). Before and after each behavioral session, we delivered trains of 5 to 10 pulses of 473 nm light, each 5 ms long, at 1, 5, 10, 20 and 50 Hz, with power
between 5 to 20 mW mm$^{-2}$. Spike shape was measured using a broadband signal (0.1–9,000 Hz) to ensure that spike waveform was not distorted.

To include a neuron in our data set, the neuron must have been well isolated (L-ratio < 0.05) (Schmitzer-Torbert et al., 2005) and recorded between two identified dopaminergic neurons or within 200 μm of an identified dopamine neuron to ensure that all neurons came from VTA. Recording sites were further verified histologically with electrolytic lesions using 15–20 s of 100 μA direct current.

**Data analysis**

To test whether the control and habenula-lesion group were different in behavioral performance during the training period, we performed a three-way ANOVA. The three factors were lesion versus control ($F_{1,338} = 2.81, P = 0.095$), trial type ($F_{3,338} = 6.19, P < 10^{-6}$) and days of training ($F_{9,338} = 128.6, P < 10^{-6}$). Same analysis was applied on data after lesion or sham-lesion operation. Both lesion operation and trial types had a significant effect on licking, but not days (lesion/sham-lesion: $F_{1,598} = 40.65, P < 10^{-4}$; trial types: $F_{3,598} = 107.44, P < 10^{-4}$; days: $F_{14,598} = 0.47, P = 0.95$).

To identify neurons as dopamine, we used a stimulus-associated spike latency test (SALT) algorithm (Kvitsiani et al., 2013) to determine whether light pulses significantly changed a neuron’s spike timing (Figure 1.4). For SALT algorithm, we used a time window of 10ms after laser onset and a significance value of $P < 0.05$. To ensure that spike sorting was not contaminated by light artifacts, all light-identified dopamine neurons had Pearson’s correlation coefficients greater than 0.9 between spontaneous and light-evoked spike waveforms, as described in (Cohen et al., 2012). These criteria together allow us to identify neurons expressing
channelrhodopsin-2 unequivocally, as shown in the bimodal distribution of Salt $P$ value (Figure 1.4 D).

To measure firing rates, peristimulus time histograms (PSTHs) were constructed using 1-ms bins. Average PSTHs in figures were smoothed with a box filter of 100 ms ($t \pm 50$ ms). Responses to specific behavioral events were calculated based on unsmoothed PSTH. In all the analyses, the baseline was calculated based on the activity during the inter-trial-interval immediately preceding odor onset of the same duration as the response time window. It should be noted that, because the baseline firing rates of dopamine neurons are elevated in lesion group, effects of lesions on task-related activities may be under- or over-estimated depending on the quantification methods used. With respect to our observation that the inhibitions caused during reward omissions is reduced by lesions, using the change from the baseline is a conservative approach. That is, with an increase in baseline firing, comparing the absolute firing rates during this period without subtracting the baseline is problematic because inhibitory responses could be underestimated in lesioned animals simply due to the elevated baseline firing rates. Because this phenotype is the main conclusion of the present study, we present the results using the change from the baseline throughout the manuscript. For the effects on inhibitory responses, our conclusions remain the same when we used the absolute firing rates. On the other hand, analysis based on the absolute firing rate underestimates lesion’s effects on excitatory responses (i.e. reward-predictive cues or rewards). Nevertheless, there is still a trend that reward responses are smaller in lesion group ($P = 0.05$ for 50% reward, $P = 0.13$ for free reward, Wilcoxon rank sum test). Overall, though, these results are consistent with our conclusions that dip during reward omission is more severely affected by lesions and that excitatory responses are partially compromised by lesions. To complement these analyses based on firing rates, we also quantified
how well dopamine neurons’ responses discriminated the different probabilities of reward on a trial-by-trial basis. For this, we calculated the auROC value of each neuron. An ROC value was calculated using spike counts obtained in the time windows as defined above.

To cluster the response profiles of all recorded VTA neurons, we first obtained a normalized response time course of each neuron in trials with 90% probability of reward by calculating a sliding-window auROC by comparing the distribution of firing rates in a 100-ms window against the distribution of baseline firing rates (900 ms before odor onset). We then performed a principal component analysis on the response time courses of neurons from the control group (from CS onset to 1s after reward delivery). The first four eigenvectors captures 87% of the variance, and the projections to those four principal components were used for $K$-means clustering to get three response types (clusters). Lesion group neurons were first projected to the same eigenvectors derived from control group and were clustered with $K$-means method later. Using this method, the phasic air puff response in cluster1 in the control group was significantly stronger than that in optogenetically-identified dopamine neurons. This indicates that cluster 1 may contain non-dopamine neurons. We observed strong air puff responses in a small fraction of non-light-identified neurons in cluster 1 (only 19 out of 121 neurons had phasic air puff response higher than 20 spikes/s, whereas the remaining 102 neurons had the average firing rate of 2.5 spikes/s). Thus the majority of neurons in cluster 1 still highly resemble the firing pattern of optogenetically-identified dopamine neurons. The decrease of air puff excitation in Cluster 1 neurons after lesion was largely due to the loss of these non-light-identified neurons with air puff response >20 spikes/s. The presence of non-canonical neurons in cluster 1 does not affect our main conclusions in the text.
In simulation using a Rescorla-Wagner model, we used anticipatory licking behavior as a behavioral readout of associative learning. This analysis only used trials in which rewards were anticipated with the probability of either 50% or 90%. We assumed no interaction between trials with different odor CSs. We also assumed a linear relationship between anticipatory licking and $V_N$ (the learned value of the odor CS in trial $N$). We updated the value of odor CS for 50% or 90% reward probability, separately, in a trial-by-trial manner using the following formula:

$$ RPE = r - V_{N-1} $$

If $RPE > 0$: $V_N = V_{N-1} + \alpha_p \cdot RPE$

If $RPE < 0$: $V_N = V_{N-1} + \alpha_N \cdot RPE$ (1)

RPE represents the reward prediction error signal from the previous ($N-1$) trial; $r$ represents reward received in $N-1$ trial (with value of $R$ or 0, depending on whether reward is delivered or not); $V_0$ is the initial value of the odor CS. Under these assumptions, the value of water reward $R$ was the same as the number of anticipatory licks when the animal fully learned the association between the odor and 100% water. The reward history was randomly generated based on the probability of reward, with 100 trials for each trial type per session. We set $V_0$ as 0 for the training data; for simulation of post-operative data, $V_0 = 0.5 \cdot R$ for 50% reward trials, and $V_0 = R$ for 90% reward trials (since during training the reward probability is 100%). We used the average lick rate for each session across animals to fit the average lick rate of a simulated session.

To fit the parameters ($R$, $\alpha_p$ and $\alpha_N$), we used a maximum likelihood fitting (“fminsearch” function in the MATLAB) to find the set of parameters that best predicted the experimental data. The likelihood of observing the data given the model was calculated each day on group averaged data by assuming a Gaussian noise with standard deviation of 1, for both 50% reward and 90%
rewarded trials. Then the likelihood was summed across all days of data and used for parameter fitting. We ran the simulation 10 times to obtain the standard deviation of fitted parameters.

To quantify the size of lesion, we linearly transformed the standard atlas (Franklin and Paxinos, 2008) to best match the corresponding Nissl-stained histological images using anatomical landmarks such as the hippocampus. We labeled brain regions based on the best-matching transformed atlas. We marked the area of lesion by identifying the areas that were destroyed or that no longer displayed Nissl-stained neuronal somata. We calculated the intersect area of lesioned tissue and different brain regions, including the medial and the lateral habenula, paraventricular thalamic nucleus, and hippocampus using a customized Matlab code.

**Immunohistochemistry**

After recording, mice were given an overdose of ketamine/medetomidine, exsanguinated with saline, perfused with paraformaldehyde, and brains were cut in 100 μm coronal sections. VTA sections were immunostained with antibodies to tyrosine hydroxylase (TH)(AB152, Millipore, Billerica, MA, USA) and secondary antibodies labelled with Cy3 (Jackson Immunoresearch). Sections were further stained with 4’,6-diamidino-2-phenylindole (DAPI) to visualize nuclei. Recording sites were identified and verified to be amid EYFP staining and TH staining in VTA. Habenula sections were stained with Nissl as described before (Cury and Uchida, 2010).
CHAPTER 2

Distributed and multi-step computation of dopamine prediction errors

This chapter is based on a manuscript I am preparing to submit along with Professor Uchida, Mitsuko Watabe-Uchida and other collaborators, entitled ‘Distributed and mixed information in monosynaptic inputs to dopamine neurons.’ Mitsuko and I collected most of the data. I analyzed the data with advice from Mitsuko and other collaborators. I wrote the manuscript together with Mitsuko and Professor Uchida.
Introduction

One of the major goals in neuroscience is to understand how the brain performs computations. Simple arithmetic-like operations are thought to underlie various brain functions. Previous studies have shown that firing rate responses of some neurons follow simple arithmetic. For instance, multiplication- or division-like modulations of neural response – often called gain control – have been observed in a plethora of neural systems including attentional and “non-classical” modulations of visual responses (Carandini and Heeger, 2012; Williford and Maunsell, 2006) and the modulation of odor responses by the overall input (Olsen et al., 2010; Uchida et al., 2013). In addition, various brain computations appear to rely on comparing two variables, requiring the brain to compute the difference between the two (Bell et al., 2008; Eshel et al., 2015; Johansen et al., 2010; Ohmae and Medina, 2015). However, how these computations are implemented in neural circuits remains elusive.

One potential scenario for these arithmetic computations is so-called variable binding (Gallistel and King, 2011; Marcus et al., 2014). A population of neurons encodes a variable X, and another population of neurons encodes another variable Y. Downstream neurons combine inputs encoding these variables, and output the result of a simple arithmetic, such as $X \div Y$ or $X - Y$. However, whether the above scenario is a plausible way to understand computations in complex neural circuits remains to be examined. First, the brain consists of neurons that are inter-connected in a highly complex manner. It is, therefore, unclear whether pure encoding of a single variable is a common or efficient way by which information is represented by neurons (Fusi et al., 2016). Second, a single neuron has to have the ability to perform precise computation, that is, its input-output function has to match a specific arithmetic. Previous modeling and experimental studies have shown that simple arithmetic computations may arise
from a wealth of nonlinear mechanisms that transform synaptic inputs into output firing at the level of single neurons (Chance et al., 2002; Holt and Koch, 1997; Silver, 2010). However, it remains unknown whether these mechanisms underlie brain computations in a natural, behavioral context.

To address these questions, here we chose to study the midbrain dopamine system. It is thought that dopamine neurons in the ventral tegmental area (VTA) compute reward prediction error (RPE), that is, actual reward minus expected reward (Bayer and Glimcher, 2005; Cohen et al., 2012; Hart et al., 2014; Schultz et al., 1997), or

\[
\text{Dopamine response} = V_{\text{actual}} - V_{\text{predicted}}
\]

(1)

where \(V_{\text{actual}}\) and \(V_{\text{predicted}}\) are the actual and predicted values of reward, respectively. The following experimental results supported this arithmetic operation. Dopamine neurons are activated by unpredicted reward. When reward is predicted by a preceding sensory cue, however, dopamine neurons’ response to reward is greatly reduced, while the reward-predictive cue starts to activate dopamine neurons (Mirenowicz and Schultz, 1994; Schultz et al., 1997). When a predicted reward is omitted, dopamine neurons reduce their firing below baseline at the time reward was expected. A recent study showed that the reduction of dopamine reward response by expectation occurs in a purely subtractive fashion (Eshel et al., 2015). Moreover, it was shown that single dopamine neurons in VTA compute RPEs in a very similar way (Eshel et al., 2016).

Although the above simple equation explains dopamine neurons’ responses to reward, it does not explain their responses to reward-predictive cues. More elaborated models of dopamine neurons use a form of temporal difference (TD) learning, a technique used to train computer programs in artificial intelligence (reinforcement learning algorithms) (Schultz et al., 1997;
Sutton, 1988b). This model computes the difference of values between two consecutive time points and takes the following form:

\[
\text{Dopamine response} = r(t) + V(t+1) - V(t)
\]  

(2)

where \( r(t) \) and \( V(t) \) denote reward obtained at time \( t \) and the value of the state at time \( t \), respectively. This simple arithmetic model successfully accounts for dopamine neurons’ responses including their response to reward-predictive cues.

Over the past twenty years, many electrophysiological recording experiments have been conducted with the goal of identifying specific brain areas conveying information about each variable in the above equations (Keiflin and Janak, 2015; Okada et al., 2009b; Schultz, 2015). Based on these data, many models have been proposed to explain RPE computations (Brown et al., 1999; Daw et al., 2006; Doya, 2000; Houk et al., 1995b; Joel et al., 2002; Kawato and Samejima, 2007; Keiflin and Janak, 2015; Kobayashi and Okada, 2007; Lee et al., 2012; Morita et al., 2012; O’Reilly et al., 2007; Schultz et al., 1997). These models typically assume that a particular brain area or a specific population of neurons within an area represents specific information such as \( V_{\text{actual}} \), \( V_{\text{predicted}} \), \( r(t) \), \( V(t+1) \) or \( V(t) \), and that these types of information are sent to dopamine neurons to compute RPEs. However, these hypotheses have not been tested experimentally. One obstacle in testing these models has been the inability to characterize both the activity of neurons and their precise synaptic connectivity (i.e., whether recorded neurons are presynaptic to dopamine neurons).

Several attempts have been made to examine neuronal activity with connectivity. Antidromic electrical stimulation has been used to identify neurons that project to a specific brain area (Hong and Hikosaka, 2014; Hong et al., 2011). More recently, optogenetic stimulation
or viral infection at axon terminals has been used in a similar manner (Jennings and Stuber, 2014; Jennings et al., 2013; Nieh et al., 2015). However, these methods cannot distinguish cell types in a projection target. Importantly, ventral tegmental area (VTA) consists of different types of neurons, not only dopaminergic but also GABAergic and others (Nair-Roberts et al., 2008; Sesack and Grace, 2009). In the present study, we sought to characterize the activity of neurons presynaptic to dopamine neurons by combining the rabies-virus-mediated trans-synaptic tracing and optogenetic-tagging with electrophysiological recording. Our results demonstrate that the information that dopamine neurons receive in order to compute RPE is not localized to specific brain areas and is not limited to pure signals of reward and expectation. Rather, information in monosynaptic inputs across multiple brain areas contained a spectrum of signals – including pure information and partial RPE – that dopamine neurons then combine into complete RPE. Our results challenge simple neural circuit models and point to the importance of understanding how neural circuits with complex connectivity can perform simple arithmetic.

Results

Identification of monosynaptic inputs to midbrain dopamine neurons during awake electrophysiology

We aimed to record the activity of neurons that are presynaptic to dopamine neurons (“input neurons”). To achieve this, we used a rabies virus-based retrograde trans-synaptic tracing system (Wertz et al., 2015; Wickersham et al., 2007a). The rabies system with mouse genetics allows us to label only neurons that directly make synapses onto dopamine neurons (Watabe-Uchida et al., 2012). Using this system, the light-gated ion channel, channelrhodopsin-2 (ChR2) (Boyden et al., 2005; Osakada et al., 2011) was expressed in input neurons so that we can
identify input neurons in behaving mice while we recorded their activity extracellularly (Cohen et al., 2012; Kvitsiani et al., 2013; Lima et al., 2009; Tian and Uchida, 2015).

**Figure 2.1 Experimental design and behavioral task**

(A) Experimental design and configuration of electrophysiological recording from rabies virus-infected neurons. (B) Probabilistic reward association task. Odor A and B predicted reward with different probabilities (90 and 50%, respectively). Odor C predicted no outcome. (C) Anticipatory licking behavior (1-2s after odor onset) after rabies injection (mean ± SEM). The dash line separated the data into early and late days in Figure 2.2. For all days plotted, the strength of anticipatory licking follows odor A > odor B > odor C (P<0.001, Wilcoxon signed rank test).

Before behavioral training, we injected helper viruses for rabies-virus tracing into the left VTA of transgenic mice expressing Cre-recombinase under the control of the dopamine transporter gene (DAT-Cre) (Bäckman et al., 2006). After recovery, the mice were then trained
in an odor-outcome association task where different odors signaled the probability of upcoming reward (90%, 50% or 0; n = 81 mice) (Figure 2.1 A). To minimize the period of rabies virus infection, we injected rabies virus (SADΔG-ChR2-mCherry(envA)) after animals acquired the association, and started recording neural activity soon after recovery from the surgery (Figure 2.1 B). We first examined the mice’s behavior during the task. Similar to the animals injected with adeno-associated virus (AAV) (Tian and Uchida, 2015), rabies-infected mice showed anticipatory licking behavior, monotonically increasing with the probability of upcoming water, over the period of 5 to 15 days after the rabies injection (Figure 2.1 C). We also observed that over the course of retraining, our mice tended to show more anticipatory licking, suggesting that odor-outcome association became stronger or mice became more motivated to perform the task during this period (Figure 2.1 C).

Figure 2.2 Recording of VTA neurons in rabies virus-infected or AAV-infected control animals.

(A) Response profiles of recorded VTA neurons in Odor A trials. VTA neurons were clustered into 3 clusters (see Materials and Methods). Each row reflects the area under the receiver-
**Figure 2.2 (Continued).** operating characteristic curve (auROC) values for a single neuron (sliding window of 100ms). Baseline is taken as one second before odor onset. Yellow, increase from baseline; cyan, decrease from baseline. Orange: neurons recorded in AAV control. Green: rabies-infected neurons. Neurons in cluster 1 with baseline firing rate less than 15 Hz were used as putative dopamine neurons in B and Figure 2.3 (B) Firing patterns of putative dopamine neurons. Average firing patterns of rabies infected putative dopamine neurons during the early (5-9 days, top) and late (10-15 days, middle) periods after rabies virus injection. Putative dopamine neurons that were infected by AAV were shown in the bottom. Only rewarded trials are shown.

We next examined whether infection of rabies virus affected neurons’ firing patterns. In this experiment, both AAV helper viruses and rabies virus were injected into VTA and recording was performed at the injection site (**Figure 2.2 B**). We compared the firing patterns of rabies-infected VTA neurons against control VTA neurons recorded in a different set of mice that were infected with AAV encoding ChR2 (Tian and Uchida, 2015). Using responses to light stimulation, we identified 76 rabies-infected VTA neurons from 12 animals. Among these neurons, rabies-infected, putative dopamine neurons showed aspects of RPE coding: (1) they were excited by reward-predictive cues with a monotonic increase in firing rate with increasing reward probability, (2) their reward responses were more strongly reduced with higher expectation, and (3) they showed inhibitory responses upon reward omission (**Figures 2.2 and 2.3**, see Methods for details on how we identify dopamine neurons). Overall, we also noticed that the strength of responses was weaker in rabies-infected putative dopamine neurons compared to those in control mice (**Figure 2.3**). The following results indicate that this difference is not
primarily due to the toxicity of rabies virus. We observed that dopamine neurons showed stronger responses to reward predictive cues in later days (10-15 days after rabies injection) compared to early days (5-9 days), contrary to the potential toxicity that would grow over longer infections (Figure 2.2 B). These results show that recorded putative dopamine neurons exhibited relatively intact RPE coding, suggesting that inputs to dopamine neurons must be sufficient to generate these responses.

**Figure 2.3 Comparison of RPE signals between rabies infected and AAV infected dopamine neurons.**

(A) Rabies group at 10-15 days after rabies injection versus AAV control. ***, P < 0.001 (Wilcoxon signed rank test against 0 spikes/s).** **P < 0.01; n.s. P >0.05 (Wilcoxon rank-sum test. Rabies versus AAV control).** (B) Percent of neurons that showed significant difference (P <
Figure 2.3 (Continued). 0.05, Wilcoxon rank-sum test) for RPE components. **$P < 0.01$, n.s. $P > 0.05$, (binomial test, between rabies versus AAV control.

Figure 2.4 Histological verification of the recording locations of input neurons.
Figure 2.4 (Continued). The recording location of each identified input neuron was marked as a circle on the schematics of coronal sections of mouse brain atlas (Franklin and Paxinos, 2008). In each column, sections are arranged from anterior to posterior. AcbSh, accumbens nucleus shell; AcbC, accumbens nucleus core; Tu, olfactory tubercle; CPu, caudate putamen (dorsal striatum); VP, ventral pallidum; LPO, lateral preoptic area; EP, entopeduncular nucleus; ZI, zona incerta; PLH, peduncular part of the lateral hypothalamus; MCLH, magnocellular nucleus of the lateral hypothalamus; PSTH, parasubthalamic nucleus; STh, subthalamic nucleus; LH, lateral hypothalamic area; RRF, retrorubral field; PnO, pontine reticular nucleus oral part; MnR, median raphe nucleus; PPTg, pedunculopontine tegmental nucleus; SPTg, subpeduncular tegmental nucleus; PB, parabrachial nucleus.

Diverse activities of monosynaptic inputs to dopamine neurons

Previous studies have mapped monosynaptic inputs to dopamine neurons using rabies virus and showed that input neurons are densely accumulated around the medial forebrain bundle (Menegas et al., 2015; Watabe-Uchida et al., 2012). We chose to record from those input-dense areas along this ventral stream that have most often been used in RPE models: the ventral and dorsal striatum (Aggarwal et al., 2012; Doya, 2008; Kawato and Samejima, 2007), the ventral pallidum (Kawato and Samejima, 2007; Keiflin and Janak, 2015; Tachibana and Hikosaka, 2012), lateral hypothalamus (Nakamura and Ono, 1986; Watabe-Uchida et al., 2012), subthalamic nucleus (Watabe-Uchida et al., 2012), rostromedial tegmental nucleus (RMTg) (Barrot et al., 2012), and pedunculopontine tegmental nucleus (PPTg) (Keiflin and Janak, 2015; Kobayashi and Okada, 2007; Pan and Hyland, 2005) (Figure 2.4). We recorded from 1,931
neurons from 69 mice, among which 205 neurons were identified as being monosynaptic inputs of dopamine neurons based on their responses to light (see Methods).

Figure 2.5 Each input area contains input neurons with diverse response profiles. (A) Configuration of input neuron recordings. (B) Examples of randomly sampled individual identified input neurons from three input areas. (C) Temporal profiles of all input neurons and VTA neurons in 90% reward trials. Color indicate an increase (yellow) or decrease (blue) from
**Figure 2.5 (Continued).** baseline, as quantified using a sliding–window auROC analysis (time bin: 100ms, against baseline). Neurons are clustered into five clusters, separated by horizontal redlines. Within each cluster, neurons are ordered by their average responses during 2 secs after odor onset. (D) Average PSTH of units from each cluster in C. n, number of units. (E) Percent of neurons belong to each of the five clusters in individual input area and VTA. VTA putative GABA, VTA other neurons are cluster 2 and 3 neurons from the AAV control group (see Figure 2.2) (F) Percent of variance explained by the first n principal components (PC) in identified inputs and dopamine neurons. PCs are computed for each individual input area and plotted as light green traces. The solid green trace is mean ± SD of traces from all input areas. Subthalamic area only has 7 neurons and is excluded from the analysis.

We first characterized each input neuron’s activity patterns in different trial types (Figure 2.5). We found that monosynaptic input neurons exhibited diverse firing patterns and complex temporal dynamics (Figure 2.5 B, 2.6). In this and the following sections, we will examine the characteristics of activities of input neurons using different forms of analysis. We first used k-means clustering to classify the activity profiles of individual neurons into five types based on a similarity metric (Pearson’s correlation coefficient) (Figure 2.5 C, see Methods). Cluster 1 showed phasic excitation to reward predictive cues and/or reward, similar to dopamine neurons (Figure 2.5 D). Clusters 2 and 3 showed sustained excitation to reward predictive cues. Cluster 4 mainly showed sustained inhibition to reward predictive cues. Cluster 5 often showed biphasic activities, first excitation followed by inhibition, to reward predictive cues. Each input area contained several different clusters of neurons, whereas all dopamine neurons fell into Cluster 1, suggesting that in contrast to dopamine neurons, input neurons in each brain area are diverse in their response profiles (Figure 2.5 E). Supporting this idea, when we extracted principal
components (PCs) of neuronal activities in dopamine neurons and each input area, the first PC accounted for 57% of firing of dopamine neurons, whereas the first PC explained only 22-34% of the activities in input neurons in each brain area (Figure 2.5 F).

Figure 2.6 Response profiles of input neurons in trials with probabilistic rewards. (A) Temporal profiles of all input neurons in 50% reward trials using a sliding-window auROC analysis. The top panels are the responses in trials in which reward was delivered. The bottom panels are responses in trials in which reward was omitted. Conventions are the same as in Fig. 2.5 C. (B) The difference in responses between rewarded trials and omission trials for each input
Figure 2.6 (Continued). neuron. The auROC value was computed at each 100 ms time bin using firing rate distributions from rewarded and omission trials in 50% probabilistic reward trials. This difference was used to identify reward responsive neurons in Figure 2.8.

Figure 2.7 Correlation analysis of dopamine neurons and their inputs

(A) Correlation matrix of firing patterns among dopamine neurons and their inputs neurons. The red lines separate neurons by brain areas. The purple ticks mark the boundaries of different
Figure 2.7 (Continued). clusters within each area (see Figure 2.5). Within each area, neurons are ordered in the same way as in Figure 2.5 C. (B) Example firing profiles of dopamine neurons and PPTg neurons (n = 7 each). Note that dopamine neurons’ firing profiles are more similar to one another than PPTg neurons. The similarity of response profiles between pairs of neurons was quantified based on the Pearson’s correlation coefficient. (C) Histogram of pairwise correlations for all pairs of neurons in each input area and dopamine neurons. The red line indicates the mean pairwise correlation for each area. The grayscale indicates the probability density.

To further quantify the diversity of response profiles, we computed the correlation coefficient of response profiles in all trial types for pairs of neurons. This analysis showed that all the dopamine neurons showed high correlation of activities to one another (indicated by a dense yellow square at top left in Figure 2.7 A), whereas input neurons in each brain area exhibited more complicated correlation patterns, often grouped as indicated by our clustering analysis (diagonal patches in Figure 2.7 A, see Figure 2.7 B for examples). When we compared the correlation coefficient of pairs of neurons within each area, dopamine neurons tended to have higher correlations than input neurons in each area (Figure 2.7 C). These analysis show that the diversity of response profiles is dramatically reduced as the information converges from the presynaptic neurons to postsynaptic dopamine neurons.

Distributed and mixed information in monosynaptic inputs to dopamine neurons

To characterize the activities of input neurons further, we next examined what information each monosynaptic input neuron conveyed to dopamine neurons with respect to the RPE computation. It should be noted that monosynaptic input neurons of dopamine neurons might also send collaterals to exert other functions through other postsynaptic targets, but our
analysis does not aim to characterize these functions. To compute RPEs in this task, dopamine neurons need to take into account at least two types of information: the value of predicted reward, which was provided by odor cues, and the value of actual reward, which was signaled by the delivery of water. The probabilistic reward task that we used (Figure 2.1A) allowed us to operationally classify the types of information that each neuron conveyed. Because the amount of reward was held constant among 90% and 50% reward trials, those neurons that conveyed pure reward information should show the same response regardless of the level of expectation. Conversely, those neurons that conveyed reward expectation should change their activity depending on the probability of reward.

Figure 2.8 Mixed coding of reward and expectation signals in input neurons.
Figure 2.8 (Continued). (A-D) Response patterns (mean ± SEM) of example neurons that were categorized as different response types. Neurons are from dorsal striatum (A), lateral hypothalamus (B), RMTg (C) and ventral striatum (D). (E) Flatmap summary (Swanson, 2000) of percent of input neurons and control VTA neurons that were classified into each response type. VTA putative GABA, VTA other neurons are cluster 2 and 3 neurons from the AAV control group (see Figure 2.2). The thickness of each line indicates the number of inputs neurons in each area (input per 10,000 total inputs in the entire brain) as defined at the top right corner (Watabe-Uchida et al., 2012). DS, dorsal striatum; VS, ventral striatum; VP, ventral pallidum; LH, lateral hypothalamus; STh, subthalamic nucleus; dopamine, dopamine neurons.

We examined whether there were neurons that conveyed actual reward values. Although previous studies found that local GABA neurons in VTA may signal reward expectation to dopamine neurons (Cohen et al., 2012; Eshel et al., 2015), these studies did not find neurons that signal actual reward in VTA. Some of the areas that we recorded are good candidates for the source of excitatory inputs for reward. For example, the lateral hypothalamus and PPTg are main sources of the few glutamatergic inputs onto dopamine neurons, whereas many other input areas of dopamine neurons are dominated by GABA neurons (Geisler and Zahm, 2005; Geisler et al., 2007; Watabe-Uchida et al., 2012). Indeed, previous recording studies found reward-responsive neurons in these areas (Nakamura and Ono, 1986; Okada et al., 2009b). We found some input neurons that appeared to convey the value of actual reward, as predicted by previous models (Figures 2.8 A, B). However, we noticed that many other neurons were modulated by reward in more complex manners (Figure 2.8 D). To further quantify this observation, we first identified reward-responsive neurons based on statistically significant differences in firing among 50% reward trials, comparing when the reward was actually delivered and when it was not (P < 0.05,
Wilcoxon rank-sum test; 2.0-2.5s from odor onset, see Figure 2.9). Although 54.6% (112 of 205) of input neurons were classified as reward responsive, a large portion (79 of 112) of these reward-responsive neurons was also modulated by the predicted value of reward ($P < 0.05$; Wilcoxon rank-sum test, 90% reward vs 50% reward, 0-500ms from reward onset) (e.g. Figure 2.8 D; Figure 2.8 E, green “mixed”). Reward responsive neurons that did not show a significant modulation by predicted value of reward can be classified as “pure reward” neurons, the type of neuron assumed in many models (e.g. Figure 2.8 A and B; Figure 2.8 E, light blue and blue). However, “pure reward” neurons often showed responses in an earlier time window, responding to odor cues (e.g. Figure 2.8 B). Removing those neurons responding to odor cues (see Figure 2.9), only 4.9% of input neurons (10 out of 205) signaled actual reward in a pure fashion (e.g. Figure 2.8 A; Figure 2.8 E, blue). These “pure reward” neurons were not localized in a specific brain area such as the lateral hypothalamus or PPTg. We found slightly more pure reward neurons in the ventral pallidum, which also contains glutamatergic projections to VTA (Geisler et al., 2007).

Similarly, we categorized “pure expectation” neurons that encoded predicted reward value monotonically and were not reward-responsive (e.g. Figure 2.8 C; Figure 2.8 E, orange; see Figure 2.9 for criteria). In addition to VTA GABA neurons (Cohen et al., 2012; Eshel et al., 2015), the striatum and ventral pallidum have been proposed to send reward expectation signals to dopamine neurons (Kawato and Samejima, 2007). However, we found that relatively few neurons (26 of 205 neurons) signaled reward expectation in a pure fashion. In fact, many putative GABA neurons in VTA were also reward responsive, although they were thought to encode reward expectation signals (Cohen et al., 2012; Eshel et al., 2015). Similar to reward neurons, expectation neurons were also distributed in several brain areas. The striatum and
ventral pallidum had slightly more expectation neurons, although these areas also had a comparable number of mixed-coding neurons.

Figure 2.9 Criteria used to categorize neurons into pure reward and pure expectation types.

(A) “Pure reward” neuron. Left, schematics of criteria used for identifying the “pure reward” type. Pure reward neurons are defined as being reward-responsive without being modulated by expectation of reward. The colors of horizontal bars encode trial types and their heights indicate
Figure 2.9 (Continued). the response magnitude as measured by the average spike counts in the specified time window. A neuron is categorized into a specific response group if its responses satisfy all of the statistical criteria illustrated. All of the \( P \) values listed are from Wilcoxon rank-sum test. Time windows used for performing statistical tests are labeled. Same conventions are used in (B-D). The example pure reward neuron on the right is from lateral hypothalamus. (B) “Pure reward no cue” neuron. A “pure reward no cue” type neurons are “pure reward” types that did not show significant modulation from baseline firing before reward was delivered. An example neuron from the dorsal striatum is shown on the right. (C) “Pure expectation” neurons, which are neurons that encode predicted reward value monotonically (90% reward > 50% reward > no reward or 90% reward < 50% reward < no reward) and are not reward-responsive. An example neuron from the RMTg is shown on the right. (D) “mixed” neurons, which are neurons that are reward-responsive but do not belong to pure reward type. An example neuron from the ventral striatum is shown on the right.

In total, these results demonstrate two characteristics of input neurons of dopamine neurons. First, each brain area that we tested contained input neurons encoding different types of information. In other words, information about the predicted and actual reward values was conveyed by input neurons spread across multiple, distributed brain areas. Second, at the single neuron level, these components of information were already mixed in most of the input neurons.

Among the neurons with mixed coding of predicted and actual reward, we found that some (17.6%, 36 out of 205) input neurons showed an important aspect of RPE coding (Figure 2.10): their reward responses were suppressed more by 90% reward expectation than by 50% reward expectation (Figure 2.10 E, pink bar; see Methods and Figure 2.11 for criteria). This type of expectation-dependent suppression of reward response was not localized in a specific
area, such as RMTg, which was proposed to send RPE to dopamine neurons (Hong et al., 2011), but were found in input neurons in all areas that we examined (Figure 2.10 E). However, RPE coding in these neurons was not as complete as it was in dopamine neurons. In addition to expectation-dependent reductions of reward responses, dopamine neurons respond to reward-predictive cues and reward omission, which are additional hallmarks of RPE signals. Only 3.9% of input neurons showed all three features of RPE signals, but many other neurons showed partial RPE signaling (Figure 2.10 E). These results demonstrate that a significant fraction of monosynaptic input neurons already conveyed aspects of RPE signals, although the quality of RPE signals is not as robust and mature as in dopamine neurons.

Figure 2.10 Input neurons encode reward prediction error signals.
Figure 2.10 (Continued). (A-D) Response patterns (mean ± SEM) of example neurons that were categorized as encoding different parts of RPE signals. Neurons are from RMTg (A, B), PPTg (C) and ventral pallidum (D). (E) Percent of input neurons and control VTA neurons that showed RPE signals.

Figure 2.11 Criteria used to categorize neurons into RPE-related response types. (A) “Partial RPE (reward)” neuron. Neurons whose reward response was modulated by reward expectation are categorized into this type. Left, a schematic of criteria used to categorize neurons
Figure 2.11 (Continued). into this group. Refer to Figure 2.10 for convention of the schematics. An example neuron from RMTg is shown on the right. (B) “Partial RPE (reward and cue)” neuron. Neurons in this type are those in (A) that also showed value coding during the cue period (0-500ms after odor onset). That is, neurons that encoded RPE-related responses during both reward and cue periods are classified into this type. An example neuron from RMTg is shown on the right. (C) “Complete RPE” neuron. Neurons that encode a complete set of RPE signals including reward omission responses are classified into this type. An example neuron from PPTg is shown on the right.

Our identification of input neurons relied on optogenetic activation. It should be noted that laser stimulation could indirectly excite neurons postsynaptic to ChR2-expressing neurons. To distinguish direct versus indirect activation, we used high-frequency light stimulation (20 and 50 Hz) because synaptic transmissions tend to fail at high frequency. To further eliminate contamination with indirect stimulation, we performed the above analysis using only neurons that responded to laser with short latency (< 6ms, 118 out of 205 neurons). The data still supported our assertion that the plurality of input neurons encoded mixed information (Figure 2.12). Furthermore, to minimize the effect of toxicity caused by rabies virus, we examined whether the results differed in early versus late days after rabies injection. Mixed coding was a dominant feature regardless of the recording period (34% and 41% neuron from 5-9 and 10-16 days after injection, Figure 2.12). These results suggested that neurons presynaptic to dopamine neurons often showed mixed coding of predicted and actual rewards, and this is unlikely due to the contamination of neurons indirectly activated by ChR2 or to toxicity of the rabies virus. We also found that mixed coding neurons and RPE neurons were more concentrated in inputs than in
other neurons in the same brain areas (Figure 2.13), indicating that our identification of inputs was not random.

Figure 2.12 Light identification criteria and rabies recording days do not affect the main results.

(A) Percent of input neurons (aggregated across all input areas) that are classified as pure reward, pure expectation, or mixed response types. “All” indicates the same population of input neurons as Figure 2.8. The middle (“Short latency < 6 ms”) shows the result using the input neurons with more strict light identification criteria (< 6 ms spike latency to the onset of laser stimulation, 118 out of 205 neurons). The right (“Early rabies 5-11 days”) shows the result using the input neurons recorded during the early half of recording days (5-11 days) after injection of rabies virus. (B) Percent of input neurons (aggregated across all input areas) that showed partial or complete RPE signals. The same population of neurons as in (A) was used in (B).
Figure 2.13 Mixed and RPE types of neurons are more concentrated in other (non-light identified) neurons.

Left, percent of neurons that showed “Pure reward”, “Pure expectation” and “mixed” response types in input neurons and other not light responsive units. **, $P < 0.01$ (Fisher's exact test).

Right, percent of neurons that showed RPE responses. ***, $P < 0.001$; **, $P < 0.01$ (Fisher's exact test).

**Contribution of negative and positive responses in input neurons**

In the previous sections, we observed that components of information for RPE computation such as predicted and actual reward values are distributed in multiple brain areas. These brain areas have different compositions of neurons; the ventral and dorsal striatum consist of inhibitory neurons (Oertel and Mugnaini, 1984), the subthalamic nucleus is excitatory (Smith and Parent, 1988), RMTg is mainly inhibitory (Jhou et al., 2009a), PPTg is mainly excitatory (Charara et al., 1996) and ventral pallidum and lateral hypothalamus are a mixture (Geisler et al., 2007; Kalivas et al., 1993; Nieh et al., 2015). To further understand which brain areas are important for generating the phasic responses of dopamine neurons upon reward-predictive cues and reward, we broke down the responses into positive and negative coding types with “positive”
type referring to those that showed a monotonic increase in firing rate response with increasing reward probabilities.

We first examined the number of neurons whose firing rates were modulated monotonically by reward probability during the CS period (0-500 ms after CS onset). We found that neurons whose firing was modulated monotonically by reward probabilities were found in all the brain areas that we examined (Figure 2.14). We also compared the results of identified input neurons with those neurons that were not identified as inputs by our optogenetic-tagging method (hereafter “unidentified neurons”). Interestingly, we found that input neurons were significantly enriched by positive reward coding compared to unidentified neurons in the same area, whereas negative reward coding was rarer both in inputs and unidentified neurons, except RMTg (Figure 2.14 A). These results suggest that the positive responses in the ventral pallidum, STh, LH and PPTg are good candidates to cause positive phasic responses of dopamine neurons in response to reward-predictive CS while the positive responses in the striatum and RMTg are not, because the later areas are known to send inhibitory inputs to dopamine neurons (Jhou et al., 2009a; Oertel and Mugnaini, 1984). Next, we compared onset latency of reward coding in inputs, other (unidentified) neurons and dopamine neurons. We plotted the percent of neurons which showed shorter latency than the median latency of dopamine neurons. Interestingly, in each area, neurons that met the two criteria – those that showed positive reward coding and short latency – were enriched in input neurons compared to unidentified neurons (Figure 2.14 B). Taken together, these results suggest that input neurons which are able to initiate value coding responses during CS in dopamine neurons were distributed in excitatory neurons in the ventral pallidum, lateral hypothalamus and PPTg.
Figure 2.14 Comparison of CS responses in identified input neurons and other neurons

(A) Percent of neurons encoding value positively or negatively during CS (0-500ms since odor onset) in identified inputs and other neurons. ***p < 0.001; **p<0.01;*p<0.05 (Fisher’s exact test). (B) Percent of neurons encoding value faster than median latency of dopamine neurons (see Methods).

On the other hand, during late delay period before reward (1.5 – 2s after CS onset), both positive and negative reward coding were distributed in multiple brain areas (Figure 2.15). A previous analysis (Figure 2.8 E, orange) indicated that a significant number of input neurons in ventral and dorsal striatum and ventral pallidum encode “pure expectation” regardless of reward. Here we found that many of these neurons encode reward positively during late delay period, suggesting that GABAergic neurons in these areas can provide expectation signals. Monotonic negative reward coding was observed 34.6% of input neuron in PPTg, suggesting a unique function of PPTg to signal reward expectation potentially via its excitatory projections (Figure 2.15).
Figure 2.15 Comparison of CS responses in identified input neurons and other neurons

Percent of neurons encoding value positively or negatively during late delay (1500-2000ms since odor onset) in identified inputs and other neurons. *p<0.05 (Fisher’s exact test).

RPE can be robustly reconstructed by a combination of monosynaptic inputs

Using the activities of all the identified input neurons, we explored whether a simple linear combination of inputs can reconstruct dopamine activities (Figure 2.16). Since the actual strength of connection of each input neuron is unknown (which may differ depending on synaptic strength, number and location of synapses etc.), the weights of each input were assigned to best fit the dopamine activities. Using the activity patterns of all of our input neurons, the model was able to fit dopamine activities with high precision ($R^2 = 0.96$). Even if one entire area is removed, fitting the dopamine activities with the other inputs still achieve high precision.
\( R^2 \geq 0.93 \) for removal of any single input area. These results support a highly redundant distribution of information across multiple brain areas.

Figure 2.16 A Linear model to fit dopamine neurons’ task responses using inputs. (A) Schematics to illustrate the linear model. The detail of model fitting is described in Methods. (B) Distribution of fitted weights. (C) The linear model predicted dopamine activity and the actual dopamine activity.

Next we examined whether RPE-related signals are more readily extracted from input neurons than from other (unidentified) neurons in the same areas. We found that for a given number of randomly picked neurons, dopamine activities were more readily reconstructed from input neurons than from other (unidentified) neurons (Figure 2.17 A). Further, we examined how much our linear reconstruction depends on the exact weights. In the above, we obtained the weights to achieve best fitting. Here, we examined how much the fit will degrade after increasing
the number of neurons whose weights were randomly swapped with one another. After shuffling the weights, we computed the similarity between the weighted sum of the activity of input neurons and the activity of dopamine neurons using a Pearson’s correlation coefficient. We found that the similarity gradually decreased but reached an asymptote at a relatively high similarity ($r \sim 0.4$) (Figure 2.17 B). Even after complete random shuffling, the sum of input activities resembled RPE signals of dopamine neurons (Figure 2.17 C). Interestingly, the decrease in similarity was significantly slower compared to the same analysis using other (unidentified) neurons (Figure 2.17B). These results suggest that the observed activity patterns of input neurons can provide not only a sufficient but also a robust substrate for constructing RPE-related activities of dopamine neurons in the sense that precise tuning of weights is not required. The comparison with other (unidentified) neurons further suggests that the activity profiles of input neurons are “pre-formatted” such that the downstream neurons can more easily generate RPE signals.

Figure 2.17. A linear model can reliably extract RPE signals from inputs.
(A) Comparison of quality of fitting using randomly sampled neurons from identified input neurons and other neurons in model fitting (mean ± SD). (B) Comparison of quality of fitting when shuffling the fitted weights using identified input neurons and other neurons (mean ± SD, 500 simulations). To shuffle the data, the weights of two randomly chosen neurons are swapped.
Figure 2.17 (Continued). “#Pairs swapped” indicates how many times this procedure is repeated for each simulation. (C) Mean ± SD of the fitted activity when the fitted weights of input neurons are completely shuffled.

Discussion

The activity of dopamine neurons can be approximated by simple arithmetic equations (Bayer and Glimcher, 2005; Cohen et al., 2012; Eshel et al., 2015, 2016; Hart et al., 2014; Schultz et al., 1997). It is, therefore, tempting to suggest simple models in which distinct populations of neurons that are presynaptic to dopamine neurons represent different terms in these equations, and these variables are combined at the level of dopamine neurons, akin to the variable binding scheme. As a result, previous experimental work has focused on finding neurons encoding these types of information in relevant brain areas (Cohen et al., 2012; Eshel et al., 2015; Keiflin and Janak, 2015; Okada et al., 2009b; Schultz, 2015). However, recording experiments typically find that these neurons are intermingled with other neurons with diverse firing patterns conveying different types of information. Thus, without confirming the assumed synaptic connectivity, it is unclear whether these models are valid. We, for the first time, addressed this issue by directly recording the activity of neurons that are presynaptic to dopamine neurons, in multiple brain areas. In stark contrast to the simple models proposed previously, we observed that information about reward and expectation are distributed and already mixed in monosynaptic inputs in multiple brain areas. Our data suggest that the RPE computation is not a one-step process combining pure information about reward and expectation in dopamine neurons. Nor do dopamine neurons receive complete RPE from a specific brain area. Instead, the prevalence of input neurons encoding both pure and mixed information, partial and
complete RPE signals appears to be a sign of redundant computations distributed in a complex neuronal network, which ultimately converge onto dopamine neurons to construct more complete RPE signals.

**Distributed and multi-step computation of RPEs**

In contrast to simple arithmetic equations for dopamine RPEs, midbrain dopamine neurons receive monosynaptic inputs from many brain areas (Geisler and Zahm, 2005; Menegas et al., 2015; Watabe-Uchida et al., 2012). This is unlikely due to the potential diversity of dopamine neurons. Previous studies have found that a subpopulation of dopamine neurons projecting to a specific target receives inputs from many brain areas (Beier et al., 2015; Lerner et al., 2015; Menegas et al., 2015). Which brain areas among these are important for RPE computation? Why do so many brain areas project to dopamine neurons? One possibility is that each brain area is specialized for specific functions, and many input neurons may not contribute for dopamine neurons to compute RPE in a simple behavioral task such as that used in the present study. In stark contrast to this prediction, 83.4% of input neurons ($P < 0.05$, ANOVA, see Methods for details) in 7 different brain areas were modulated by reward-predictive cues or reward. We further found that each of the two critical components of RPE computations – information about predicted and actual reward – were distributed in inputs in these brain areas.

Our data also showed that the two types of information are often mixed in single input neurons. Very few neurons represented the information about expected reward and actual reward in a pure fashion. Why is the information mixed? First, anatomical constraints may make coding of pure information unrealistic. The brain is a dynamical system consisting of a large number of neurons that are interconnected in a highly complex manner within and across brain areas.
Indeed, all the areas that we recorded in this study are highly connected with one another, either directly or indirectly.

Second, it has been proposed that mixed coding – single neurons representing multiple types of information in a mixed fashion – may garner computational and representational advantages (Fusi et al., 2016). For instance, mixing information in non-linear way may expand the space of information coding, making the read out of the downstream neurons easier (Fusi et al., 2016; Rigotti et al., 2013). This line of work suggests that the brain exploits these properties to actively mix information at the single neuron level, and mixed coding is a normal way by which the brain represents information, rather than exception. In this framework, it can be seen that particular variables and computations are embedded in a complex pattern of activity distributed in a network of many neurons.

Third, mixing information can be seen as a reflection of gradual RPE computations. It has been reported that RPE signals can be found in multiple brain areas such as prefrontal cortex, the striatum, GPi, lateral habenula and RMTg (Asaad and Eskandar, 2011; Bromberg-Martin and Hikosaka, 2011; Hong and Hikosaka, 2008; Matsumoto and Hikosaka, 2007, 2008; Oyama et al., 2015; Sul et al., 2010). This is likely because RPE signals are correlated with various behaviors in addition to reinforcement learning (Den Ouden et al., 2012; Schultz and Dickinson, 2000). Our experiment revealed, for the first time, that these RPE-coding neurons actually provide monosynaptic inputs to dopamine neurons, albeit that RPE signals in inputs were not complete, compared to those in dopamine neurons. Peculiarly, these RPE-coding neurons were more enriched in input neurons compared to other neurons in the same brain area. Furthermore, our simulation showed that reconstruction of RPE signals was easier using the activity of input neurons compared to that of other (unidentified) neurons. The enrichment of mixed coding in
input neurons thus appears to be a reflection of the gradual step toward RPE computation. In parallel, some input neurons send pure information of predicted and actual reward values, potentially further reinforcing the fidelity of RPE computations. In total, RPEs appear to form in a multi-layered, distributed network that eventually converges onto dopamine neurons.

**Comparison to past models**

Although dopamine neurons receive a significant number of glutamatergic projections, the major projections to dopamine neurons are GABAergic (Tepper and Lee, 2007). Because of this anatomical observation, many models proposed that phasic excitation of dopamine neurons are caused by disinhibition of GABA neurons (Gale and Perkel, 2010; Lobb et al., 2010, 2011). More specifically, the basal ganglia have been proposed to provide a disinhibition circuit. Indeed, dopamine neurons receive large numbers of monosynaptic input from both the striatum and pallidum (Watabe-Uchida et al., 2012). In the present study, however, we observed that input neurons in most of the areas including the ventral and dorsal striatum and ventral pallidum were excited by positive values (e.g. reward-predictive cue). In our dataset of the striatum or ventral pallidum (n = 58 and 18 input neurons, respectively), we found no neuron that was inhibited in a monotonically value-dependent manner by reward-predictive CS in a short latency. These results suggest that the phasic excitation of dopamine neurons is likely caused by excitatory inputs rather than disinhibitions, although we cannot exclude the possibility that other brain areas from which we did not record may provide a disinhibition circuit.

If GABA neurons in inputs may not provide phasic excitation of dopamine neurons, what is the function of huge amounts of inhibitory inputs? In previous studies, we found that GABA neurons in VTA show activities which correlate with expected reward values. Furthermore, inhibition of these neurons increased phasic responses to rewards in dopamine neurons (Cohen et
al., 2012; Eshel et al., 2015). These observations suggested that VTA GABA neurons are a source of inhibitory inputs that suppresses dopamine reward response depending on reward expectation. In the present study, we observed similar activities in input neurons in each area (Cluster 2 and 3 in Figure 2.5 C, E, monotonic value coding in Figure 2.15), especially in the ventral and dorsal striatum and ventral pallidum during US period (“pure expectation” in Figure 2.8 E). These results further support the present finding of redundancy in inputs. Yet, because of the vicinity of local inputs and weak electrical connections from the striatum to dopamine neurons (Bocklisch et al., 2013; Chuhma et al., 2011; Xia et al., 2011), it is possible that local GABA neurons have a larger influence on dopamine neurons. It is important to examine the difference of synaptic weights and functions between local and long distance inputs.

In addition to providing expectation signals, inhibitory inputs may have other functions. We noticed two related characteristics in the activity of input neurons. First, input neurons tended to show more sustained activities than dopamine neurons, although a relatively small fraction of neurons showed transient phasic activities. It is possible that inhibitory inputs may contribute to make the dopamine activities more phasic than sum of excitatory inputs, akin to taking the derivative of otherwise sustained inputs. Note that taking the derivative is an important component of temporal difference error calculation (Schultz et al., 1997; Uchida, 2014). Second, many input neurons showed stronger excitation at air puff than dopamine neurons. Because inputs in the striatum, pallidum and RMTg as well as GABA neurons in VTA also showed excitation to air puff, these inputs may function to cancel out excitatory inputs that are not encoding values. Air puff response of dopamine neurons may depend on a balance of impinging, massive excitatory and inhibitory inputs.

**Technical considerations**
In our previous recordings and the present study, we have focused on dopamine neurons in the lateral VTA. In this location, most dopamine neurons showed a uniform response function for RPE computations (Eshel et al., 2016). However, increasing studies indicate the diversity of midbrain dopamine neurons. For example, putative dopamine neurons in the lateral SNC in monkeys were excited by aversive stimuli (salience type) whereas putative dopamine neurons in VTA and medial SNC were inhibited (value type) (Matsumoto and Hikosaka, 2009). Other studies also suggested that responses to aversive stimuli in dopamine neurons are variable depending on their projection targets (Lammel et al., 2014; Lerner et al., 2015). To study the common features of the majority of dopamine neurons on RPE computation, our analyses focused on reward related responses. To address how different inputs contribute to the diversity of dopamine responses, it is important to distinguish different subsets of dopamine neurons which project to different brain areas. By incorporating newly developed rabies methods with retrograde infection of helper CAV or AAV (Beier et al., 2015; Menegas et al., 2015) into the present recording system, it will be possible to record from inputs to a subset of dopamine neurons which project to a specific brain area.

A previous study showed that cortical pyramidal neurons infected with SADΔG virus showed electrophysiological properties similar to uninfected neurons 5-12 days after injection, and the survival was roughly constant until 16 days, although the survival of neurons dropped substantially 18 days after injection (Wickersham et al., 2007b). We thus restricted our recording time window to 5-16 days after injection of the rabies virus. This meant that we had only 12 days at maximum of recording per animal. This was the biggest obstacle in the present study because it was very difficult to achieve optimal behaviors of mice (Figure 2.1 C) and precise targeting of recording electrodes in this short time window. A recent study found that a different line of
rabies virus may lengthen the survival of infected neurons and have more efficient trans-synaptic spreads (Reardon et al., 2016). Improving the methodology using an improved version of rabies virus such as this will aid greatly in advancing our ability to study neural computations.

**Conclusions**

Despite these difficulties, our data appears to indicate that a simple arithmetic – the computation of reward prediction error – does not arise from a simple variable binding scheme. Instead, our data show that components of information required for RPE computation are distributed and mixed in monosynaptic inputs in multiple brain areas. As a result, simple arithmetic equations are not apparent from observing the activity of neurons that provide inputs to dopamine neurons. It is remarkable that, despite this complexity at the level of inputs, once these inputs are combined, the activity of dopamine neurons becomes homogeneous and largely follows simple arithmetic equations (Bayer and Glimcher, 2005; Eshel et al., 2015, 2016; Schultz et al., 1997). These results, at a first glance, might appear to indicate that dopamine neurons have the ability to weigh different inputs precisely to compute RPEs. Unfortunately, our methods did not allow us to measure the precise weights of each of these input neurons but our analysis also indicated that precise tuning of synaptic weights is not as important as one might imagine. We propose that the mixed and distributed representation of information in the brain network may help dopamine neurons to robustly combine inputs to converge on a single solution.

**Methods**

**Animals**

We used 99 adult male mice heterozygous for Cre recombinase under the control of the DAT gene (B6.SJL-Slc6a3tm1.1(cre) Bkmn/J, The Jackson Laboratory) (Bäckman et al., 2006),
backcrossed for more than 5 generations with C57/BL6C mice. Animals’ age ranged from 2 to 9 months. 81 mice contributed to the recording dataset. The other animals were not able to finish the whole experiment or mistargeted, and therefore were excluded from the analyses. Animals were housed on a 12 h dark/12 h light cycle (dark from 7:00 to 19:00) and performed the task at the same time each day. All procedures were approved by the Harvard University Institutional Animal Care and Use Committee.

**Surgery and viral injections**

All surgeries were performed under aseptic conditions with animals under either ketamine/medetomidine (60 and 0.5 mg kg\(^{-1}\), intraperitoneal, respectively) or isoflurane (1–2% at 0.5–1.0 l min\(^{-1}\)) anesthesia. For analgesia, buprenorphine was administered right before the surgery (0.1 mg kg\(^{-1}\), intraperitoneally) and ketoprofen was injected postoperatively (5 mg kg\(^{-1}\) intraperitoneally). For the recording experiments, mice underwent two surgeries. In the first surgery, we used a dental cement to implant a customized metal head plate onto mice’s skull. During this surgery, we injected 500–700 nl of AAV5-FLEX-TVA (4 x 10\(^{12}\) particles/ml) and AAV8-FLEX-RG (2 x 10\(^{12}\) particles/ml) into VTA stereotactically to enable expression of TVA and RG in dopamine neurons (Watabe-Uchida et al., 2012). We targeted left VTA (3.1 mm posterior to Bregma, 0.6 mm lateral) and injected AAVs in two sites (4.0 mm and 4.2 mm ventral to dura). After training mice in the classical conditioning tasks, we performed a second surgery to inject 500–700 nl of SAD\(\Delta G\)-ChR2-mCherry(envA) (1.0 x 10\(^{7}\) plaque-forming units [pfu]/ml) (Watabe-Uchida et al., 2012) in VTA at the same stereotactic location as the first injection. We then implanted a microdrive containing six to eight tetrodes and an optical fibre either in the VTA or in one of the ipsilateral input areas. After recordings, animals were perfused
with 4% paraformaldehyde and brain sections were examined histologically to verify recording sites (**Figure 2.4**).

**Behavioral tasks**

Training of the animals was described in detail previously (Tian and Uchida, 2015). Briefly, animals were first habituated for 1-2 days, and were trained to lick the water spout upon water delivery. Mice were then trained to associate different odors with different probabilities of outcomes: 100% water, 50% water, nothing, or 100% air puff. We introduced 90% water, 80% air puff, free water and free air puff (water or air puff not preceded by any odor cues) in the last two of 10 days of conditioning training, and in recording sessions. For each animal, four odors were assigned randomly from a set of odors (p-Cymene, ethyl butryrate, isoamyl acetate, hexanol, 1-3 dimethoxy benzene, (+) carvone and butanol). Each odor cue was used in 22.5% of trials and the free water or free air puff was delivered in 5% of trials. In 24% of the identified rabies neurons, odor C was associated with 10% reward instead of nothing and we only used the unrewarded trials for data analysis. In 17% neurons of ventral pallidum and 4% neurons of dorsal striatum, odor A was associated with 100% reward instead of 90% reward. Inter-trial intervals were drawn from an exponential distribution with a minimum at 3.1 s, mean at 7.6 s and a cut off of the long delay at 20 s. Different trial types were pseudorandomized in a 100-trial block. To focus on RPE computation with reward, we didn’t use data from the air puff trials in the present study. Anticipatory licks were observed to confirm animal’s association learning. Licks were detected by breaks of an infrared beam placed in front of the water tube and the signals were digitized and recorded at 1 kHz (PCI-6251; National Instruments).

**Electrophysiology**
Recording techniques were similar to those described in previous studies (Cohen et al., 2012; Tian and Uchida, 2015). Briefly, we recorded extracellular signals using a custom-built, screw-driven microdrive containing six or eight tetrodes. Tetrodes were glued to a 200 µm optic fibre (ThorLabs) so that their tips extended 300–600 µm from the end of the fibre. Neural and behavioural signals were recorded with a DigiLynx recording system (Neuralynx) or a custom-built system using a multi-channel amplifier chip (RHA2116, Intan Technologies) and data acquisition device (PCIe-6351, National Instruments). Broadband signals from each wire were filtered between 0.1 and 9,000 Hz and recorded continuously at 32 kHz. To extract timing of spikes, signals were band-pass-filtered between 300 and 6,000 Hz and sorted offline using MClust-3.5 (A. D. Redish). At the end of each session, the fibre and tetrodes were lowered by 80 µm to record new units the next day.

Only well-isolated neurons (L-ratio < 0.05) were included in our dataset. To identify monosynaptic inputs of dopamine neurons, we used ChR2 to observe the stimulation-locked spikes (Kvitsiani et al., 2013). The optical fibre was coupled with a diode-pumped laser with analogue amplitude modulation (Laserglow Technologies and Optoengine). At the beginning and end of each recording session, we delivered trains of ten blue (473 nm) light pulses, each 5 ms long, at both low frequency (1 and 5 Hz) and high frequency (10, 20 and 50 Hz) with power between 1–20 mW mm⁻² at the tip of the fibre. Spike waveform was quantified based on a broadband signal (0.1–9,000 Hz) sampled at 32 kHz.

**Data analysis**

To identify neurons as monosynaptic inputs of dopamine neurons, we used a stimulus-associated spike latency test (SALT) algorithm (Kvitsiani et al., 2013) to determine whether light pulses significantly changed a neuron’s spike timing. For SALT algorithm, we used a time
window of 15 ms after laser onset and performed the statistical test on both low frequency and high frequency laser stimulation. Occasionally, neurons that were inhibited by laser stimulation were also detected by SALT and those neurons were not included as light identified inputs. The laser pulses could indirectly excite postsynaptic neurons of ChR2-expressing neurons, in which case the neurons should follow high frequency stimulation with much lower fidelity. To minimize contamination of indirectly activated neurons, an identified input neuron had to show significant responses to both low frequency stimulation (1, 5 Hz) and high frequency stimulation (10, 20, 50 Hz) (SALT, \( P \) value < 0.01) (Guyenet and Aghajanian, 1978). For striatal neurons, they responded to ChR2 stimulation with longer latency and less fidelity, probably due to their biophysical properties (Kravitz et al., 2013). Therefore, we only required significant modulations with stimulations at low frequencies for striatal neuron. To ensure that spike-sorting was not contaminated by light artifacts, all light-identified neurons had a high similarity between spontaneous and light-evoked spike waveforms in all tetrode wires that showed signals (Pearson’s correlation coefficients greater than 0.9). Latency of responding to light activation is also an informative indicator of whether a neuron is directly activated by light, with shorter latency more indicative of direct activation of a ChR2-expressing neuron by light. We confirmed that neurons with short latency (< 6 ms) showed consistent results compared to our main results.

To make sure that expression of rabies virus did not change the neurons’ firing patterns dramatically, we compared VTA neurons’ activities with a control data set. The control data set was collected from VTA in DAT-Cre mice which was injected with AAV5-DIO-ChR2 (Boyden et al., 2005) and published previously (Tian and Uchida, 2015). A tetrode bundle with optical fibre was implanted in VTA. Neurons were recorded around two weeks after animals acquired odor-outcome association in the same conditioning task as described above. We identified
dopamine neurons based on their response to laser stimulation (Cohen et al., 2012; Kvitsiani et al., 2013). A neuron was included as a VTA neuron if it was recorded between two identified dopamine neurons or within 200 µm of an identified dopamine neuron. Those identified dopamine neurons were used as dopamine neurons throughout the paper. For VTA neurons in the rabies group, they were identified as rabies infected neurons based on their responses to laser stimulation (SALT $P$ value $< 0.01$, wave correlation $>0.9$). We verified that the recording sites were in VTA post-mortem. To compare putative dopamine neurons’ responses between control and rabies groups, we clustered neurons’ responses into three types using VTA neurons from rabies group and AAV group (Figure 2.4). We first normalized each neuron’s response by calculating a sliding-window area under the receiver operating characteristics curve (auROC), by comparing the distribution of firing rates in a 100 ms window against the distribution of baseline firing rates (0-1000ms before odor onset, each sample measured in a 100 ms bin). We then used a k-means clustering method to cluster neurons into three types based on their responses in 90% reward trials (0-3 s from odor onset) using Euclidean distance. We previously showed that VTA neurons can be clustered into 3 clusters based on their response profiles: type 2 with sustained excitation, type 3 with sustained inhibition and type 1 for others (Cohen et al., 2012). Although our previous studies showed that all dopamine neurons fell into cluster 1 (Cohen et al., 2012; Tian and Uchida, 2015), other cell types could also be included in cluster 1. To further restrict cluster 1 neurons to putative dopamine neurons, those cluster 1 neurons that had high baseline firing ($< 15$ spikes/s) were excluded from putative dopamine neurons because none of the optogenetically-identified dopamine neurons showed baseline firing rates higher than 15 spikes/s. Applying this criterion, 85% of cluster 1 neurons from the rabies group and 75% of cluster 1 neurons in the control group were used as putative dopamine neurons. For each neuron, we
quantified the amplitude of responses during odor presentations, reward and reward omission, using a 0-500 ms time window because most firing modulations occurred within this time window. We performed Wilcoxon rank sum tests to compare the amplitudes of response or binomial tests to compare the percentage of neurons showing a significant response to specific events.

To cluster input neurons and VTA neurons, we first normalized each neuron’s response using the aforementioned auROC. Then we concatenated the normalized responses in 90% reward, 50% reward and 0% reward trials for each unit. We used a k-means clustering algorithm to cluster neurons into five clusters based on correlation similarity metrics. The number of clusters were chosen to separate distinct response types.

To compute the principal components (Figure 2.5 F), we first normalized each unit by its maximum firing rate and then subtracted by its mean firing rate. We then used the pca function in MATLAB to compute the principal components. We used neurons’ responses in 90% reward, 50% reward and 0% reward trials for this analysis.

We computed the similarity (correlations) between the response profiles of pairs of input neurons as follows. For each input neuron, we first computed the normalized responses using a sliding-window auROC (0-3 s after odor onset) in five different trial types. We then concatenated them into a single response vector. The trial types used were 90% reward, 50% reward, omission of 90% reward, omission of 50% reward and no outcome (nothing). For each input area, we iterated through all unique pairs of neurons and computed the Pearson’s correlation coefficient based on the activity vector constructed as above. Some neurons in ventral pallidum (31% of all pairs) and dorsal striatum (9% of all pairs) which had 100% reward trials instead of 90% reward trials were omitted for the calculation.
The criteria for classifying neurons into pure reward, pure expectation and mixed coding were discussed in detail in the main text and were illustrated in Figure 2.9. Here we emphasize that our approach was biased to report fewer mixed coding neurons. For example, one of the criteria for a pure expectation neuron was not to be significantly modulated by reward using a statistical test. Therefore, a pure expectation neuron identified using this criterion could still show a weak response to reward. Therefore, our conclusions that most input neurons showed mixed coding of reward and expectation hold true even under strict criteria.

We classified whether a neuron encoded RPE signals based on its responses to both cue and reward (see Figure 2.11). We used a time window of 500 ms for analysis of all events since it covered the main response time window of dopamine neurons. Specifically, “Partial RPE (reward)” neurons were (1) those that showed a significant difference between responses to 50% reward and 90% reward ($P < 0.05$, Wilcoxon rank sum test), (2) those that responded significantly different between 50% reward and 0% reward, and (3) those that the response to 90% reward was in-between the responses to 50% reward and 0% reward. “Partial RPE (reward and cue)” neurons were “Partial RPE (reward)” neurons that showed value coding during CS (see the following for details). “Complete RPE” neurons were “Partial RPE (reward and cue)” neurons which also showed significant responses to omission of 50% reward ($P < 0.05$, Wilcoxon rank sum test) and the direction of omission response was opposite to that of the other RPE responses.

We classified a neuron to encode value during CS period if there was a significant difference between the responses to odor cues predicting 90% reward and 0% reward ($P < 0.05$, Wilcoxon rank sum test) and response to 50% reward cue was in between responses to 90% and 0% reward cues. Positive value coding neurons responded stronger to 90% than 0% reward cues; conversely, negative value coding neurons responded stronger to 0% reward cues.

To compute
the latency of neurons showing value coding, a unit is test for significant difference between 90% and 0% reward cues at sliding time bin of 50ms with a step of 5ms. The latency is the center of the first time bin when five consecutive time bins all showed significance. Using this method, the median latency of dopamine neurons’ response show value coding during CS was 120ms. Similar method was applied to the analysis of free water, except that the significance test is between free reward and pre-reward baseline activities. Because response to free water is very fast, we used a 30ms time bin instead of 50ms. Using this method, the median latency of dopamine neurons’ response to free water was 115ms.

We built a constrained regularized linear regression to fit the activities of input neurons to the activity of dopamine neurons. We used responses in seven trial types with time bin of 50 ms for the model fitting (Figure 2.16 and 2.17). The cost function takes the form,

$$\|w^TX - y\|^2 + \alpha \|w\|^2$$

Where the w is the weights and X is the response vectors for each input neurons. The second term is to regularize the complexity of the model. We performed cross-validation by using every other data points as the training data set and the rest as the test data set and found $\alpha = 0.5$ gives the better fit in the test dataset. We then used $\alpha = 0.5$ for all of the fitting. The weights of individual neurons are tuned to minimize the cost function with the constraints that neurons in striatal areas can only take negative weights and subthalamic neurons can only take positive weights. To compare inputs and other neurons in model fitting with shuffled weights in Figuer 7E, we first sampled the same number of neurons as inputs out of the other unidentified neurons. We generated several sets of samples and chose the one that was most similar to total population in terms of ratios of neurons showing mixed responses, RPE responses as well as CS value responses.
To check percent of neurons that are significantly modulated by the task, we used four different time windows after odor onset: 0-500ms (CS), 1000-1500ms (early delay), 1500-2000ms (late delay), 2000-2500ms (US). For CS and delay time windows, ANOVA is performed across three trial types (90%, 50% and 0% reward CSs); for US time windows, ANOVA is performed with two additional conditions (omission of 90% reward, omission of 50% reward). A neuron shows significant task-related response if it shows significant responses in any of these four time windows, with Bonferroni correction of multiple comparisons.

**Immunohistochemistry**

On the day 16 after injection of rabies virus, recording locations were marked by making electrolytic lesions. Under anesthesia, 5-15 s of 30 µA direct current was passed through the tetrodes from which optogenetically-identified neurons were recorded. The mice were then perfused with PBS followed by 4% paraformaldehyde in PBS. The brains were cut in 100-µm coronal sections using a vibratome (Leica). Immunohistochemistry was performed using the anti-tyrosine hydroxylase antibody (AB152; Millipore, Billerica, MA, USA), and Alexa Fluoro 488 goat anti-rabbit secondary antibodies (Molecular Probes), counterstained with DAPI (Vectashield). The locations of recording sites were determined using the standard mouse brain atlas (Franklin and Paxinos, 2008). The recording depth was estimated by subtracting daily advancements of each tetrode.
CONCLUSION
Learning to approach reward and avoid punishment is fundamental to animals’ survival. Pioneering work by Kamin on animals’ learning behaviors showed that merely pairing a sensory cue with reward is not sufficient to form associations. For new associations to form, the reward has to be somewhat surprising to the animals (Kamin, 1969). Later, Rescorla and Wagner summarized those observations with a mathematical model, and suggested that the increase of association strength after each trial is proportional to the discrepancy between actual reward and expected reward (Rescorla and Wagner, 1972). The Rescorla-Wagner model explains how learning develops on a trial-by-trial basis very well. The major limitation of the Rescorla-Wagner model is that it does not account for the timing structure of inter-stimuli-interval and inter-trial-interval. To overcome this limitation, Sutton and Barto proposed the temporal difference model that can compute the value functions on a moment-by-moment basis and can model associative learning behavior with various types of inter-stimuli-intervals (Sutton and Barto, 1990). Do the neural mechanisms of reinforcement learning share any similarities with these models? Schultz and his colleagues discovered that midbrain dopamine neurons encode reward prediction error signals (Schultz, 2006; Schultz et al., 1997), which are the key signals in both the temporal difference model and the Rescorla-Wagner model.

Despite the important role of dopamine neurons in motivated behavior, it is not yet known how dopamine neurons compute reward prediction error signals. In this dissertation, I tried to tackle this question from two different perspectives. In Chapter 1, I tested a specific hypothesis that the lateral habenula provides the prediction error signals to dopamine neurons. In fact, lesion of the habenula impaired the reward omission responses in dopamine neurons relatively specifically; many other aspects of dopamine prediction error responses largely remained. To further discover the sources of prediction error signals, in Chapter 2 I took a
systematic screening approach to examine seven input areas’ responses using the same behavioral paradigm. By combining rabies virus with optogenetic identification of cell types, I was able to record the task-related activity of neurons that are monosynaptic inputs of dopamine neurons. Interestingly, information about reward expectation and reward is distributed across multiple brain areas, and often multiplexed at the single neuron level. In the following, I will discuss how these results have changed my thinking about neural circuits.

**Distributed signals vs. Specific effect of inactivation**

In Chapter 2, we showed that information is distributed across multiple input areas, yet in Chapter 1, lesioning a specific area such as the habenula impaired a relatively specific component of RPE signals in dopamine neurons (i.e., reward omission signals). Due to the redundancy of information encoded in the input areas, we might expect to see little or mild effect on dopamine firing when inactivating any individual input area. Here I provide two potential mechanisms to reconcile this issue.

First, some of the redundant information we observe in many input areas might come from a common source. Anatomically, LHb is not a strong direct input of dopamine neurons (Menegas et al., 2015; Watabe-Uchida et al., 2012), but it projects strongly to many of the input areas of dopamine neurons (Quina et al., 2015) (**Figure 3.1**). Based on the anatomical connections, it seems that LHb is able to broadcast information about reward omission to several areas then get converged onto dopamine neurons. This type of “diverge” and then “converge” propagation of information might create the seemingly redundant information distributed in many input areas. It is then possible to reveal the key area for providing such information, which may be concealed from recording results, by inactivating the source or the bottleneck area for reward omission.
Figure 3.1 Lateral habenula projects to input areas of dopamine neurons

The output fibers of the habenula form a fiber tract called the fasciculus retroflexus (fr). The fr travels ventrocadually to the midbrain and branches out in a rostral direction to the posterior hypothalamus (PH) and caudally to areas such as the median raphe (MnR), RMTg, dorsal raphe (DR), periaqueductal grey (PAG) and lateraldorsal tegmental area (LDTg). Those efferent areas of the lateral habenula often project to the VTA. Orange lines, the efferent of the habenula, based on (Quina et al., 2015). Red lines, projection to dopamine neurons, based on (Watabe-Uchida et al., 2012). Only areas relevant for both habenula and VTA dopamine neurons are shown.

Another possibility is that even though a specific signal (e.g., reward omission) is observed in many input areas, due to the biophysical properties, only one pathway is crucial for the responses in the output neurons. For example, depending on the subcellular localization of GABAergic synapses, they could have subtractive (shunting) or divisive effects on the firing of the post-synaptic neuron (Holt and Koch, 1997; Silver, 2010). To create the reward omission response in dopamine neurons, the inhibitory inputs should be able to bring down dopamine
neurons’ responses below the baseline firing rate, which requires a shunting type of inhibition. It has been proposed that GABAergic inputs from the RMTg might be suitable for shunting dopamine neurons since it preferentially localized on the proximal dendrites (Omelchenko et al., 2009).

Many questions remain to be resolved before we can fully understand how reward omission signals are generated and signaled to dopamine neurons. First, it is not yet known where the signals in the habenula originate. It has been shown that the globus pallidus internal segment in primates sends reward-related signals to the lateral habenula (Hong and Hikosaka, 2008). However, the habenula also receives inputs from many other brain regions (Aghajanian and Wang, 1977; Herkenham and Nauta, 1977). Inactivation of one input area of the habenula at a time will reveal which inputs are critical for the reward omission responses or other responses in the habenula.

Second, it will be interesting to compare the reward omission signals in the habenula and those in dopamine neurons. As mentioned before, the habenula is likely to influence dopamine neurons’ responses by passing the signals indirectly through other nuclei. Therefore, the reward omission signals from the habenula will not only be filtered by the biophysical properties of the synapses, but also be further processed by mixing with other signals in those areas. It is not known whether lateral habenula neurons can encode the amplitude of reward omission as precisely as dopamine neurons. This can be tested by including omission trials with different expected reward probabilities or reward sizes. In addition, by varying the interval between reward cue and reward, one can study whether the recorded reward omission responses correspond to the hazard rate or the reward probability density function. Moreover, the timescale on which habenula or dopamine neurons update their reward expectation signals during
omission trials is nearly unknown. A recent study showed that some dopamine neurons can encode the trained value of the visual object even when rewards are no longer contingent on the visual object while other dopamine neurons showed little response (Kim et al., 2015). It will be interesting to compare how quickly the habenula updates its value representation after changes of reward contingencies. Together, these manipulations of behavioral contexts will tell us how information about reward omission is transformed from habenula to dopamine neurons.

**Comparison to other prediction error signals**

Prediction error signals are widely used and encoded in the brain. In this section, I attempt to compare different forms of prediction error signals. I will briefly discuss sensory prediction error signals in the visual area and sensory-motor prediction error signals in the cerebellum. Although there are more unknowns than knowns in our knowledge about each of these circuits, I hope those comparisons will help us form a better understanding of the general principles as well as the specific constraints of each type of prediction error computation, and inspire new directions for future studies.

Perceptual prediction error signals have been widely observed in visual cortices when an image of a trained sequence of visual stimuli violates the subject’s expectation (Den Ouden et al., 2012). In contrast to the RPE, which only signals the amplitude and sign of the error of reward, perceptual prediction error signals often carry specific information about how the stimulus perceived is different from the prediction (e.g., V1 neuron activated by the unexpected presence of an oriented edge in its receptive field). It has been proposed that predictions are generated at the agranular layers by a top-down mechanism. Neurons encoding predicted stimuli will putatively send inhibitory projections to layer 4 neurons to inhibit the excitatory inputs about actual stimulus and create prediction error signals (Friston, 2005; Rao and Ballard, 1999;
Spratling, 2008). Although this is an attractive idea, there is as yet little evidence to show that distinct populations of neurons encode predictions and prediction errors in the visual cortices. Moreover, to compute prediction errors in a single neuron, the output neuron has to receive separate inputs about actual stimulus and predicted stimulus from similar receptive fields. It remains to be seen whether this simple model is actually implemented by the brain or not. Notably, prediction can also boost the response to expected stimuli perhaps due to the increase of attention (Gavornik and Bear, 2014). Therefore, the cortex needs additional circuits to switch flexibly between enhancing and suppressing predicted stimuli based in the behavioral context.

Another classical neural circuit that appears to perform an error-like computation is the climbing fiber inputs of Purkinje neurons in the cerebellum. The climbing fibers send error signals to activate Purkinje neurons when a novel load is used for wrist movements (Gilbert and Thach, 1977), or when the gain of visual feedback to guide movements has been changed (Ojakangas and Ebner, 1992). The excitatory inputs from climbing fibers can trigger Purkinje neurons to fire in a specific mode called complex spikes (Thach, 1967). Moreover, complex spikes in Purkinje neurons can cause long-term depression of the mossy-fiber synapses of Purkinje cells, which is thought to reduce the error signals in future movements and serve as a mechanism for motor adaptation or motor learning (Ito, 2000). If climbing fiber inputs act as teaching signals, removing this input should cause deficits in learning. Indeed, when the inferior olive nucleus, the source of climbing fibers, is lesioned, animals are no longer able to make long-term adaption of gains in the vestibular-ocular reflex, while visual stimuli are still able to modulate the amplitude of the reflex (Ito and Miyashita, 1975).

Climbing fibers have also been shown to encode prediction error during aversive conditioning. Climbing fibers are activated by aversive air puffs. After pairing the air puff with
an auditory cue, the response to the air puff is greatly reduced (Kim et al., 1998; Sears and Steinmetz, 1991). Lesion of the inferior olivary complex prevents the animals from learning the association between cue and air puff (McCormick et al., 1985). Furthermore, blocking the inhibitory input of the inferior olive nucleus prevents the animals from extinction of learned conditioned eye-blink, while blocking the excitatory input causes extinction even if the US is present (Medina et al., 2002). These results indicate that climbing fiber activities have bidirectional control over the conditioning behavior, with a positive activity enhancing association and negative activity weakening association.

How is prediction error generated in the inferior olive nucleus? This nucleus receives excitatory sensory inputs from the spinal cord and inhibitory inputs from deep nuclei in the cerebellum (De Zeeuw et al., 1998). After conditioning, climbing fibers are no longer activated by predicted air puffs, and therefore Purkinje cells do not show complex spikes. However, with infusions of the GABA antagonist picrotoxin in the inferior olive nucleus, Purkinje cells respond to predicted air puffs with complex spikes (Kim et al., 1998). Therefore, it has been proposed that the inferior olive nucleus subtracts the inhibitory feedback about the expected air puff from the cerebellum from the excitatory inputs about sensory information to compute prediction error signals. Although interesting, this hypothesis has not been directly tested by recording single unit activity in neurons presynaptic to the inferior olive nucleus.

In addition to signaling error signals for learning, the complex spikes in Purkinje cells might also be important for online control of movement. In a task in which monkeys must make rapid reaches to visual targets, the complex spikes in Purkinje cells carry information about the motor destination immediately after target onset (Kitazawa et al., 1998). After the movement, the complex spikes mainly carry information about the error signals between the motor destination
and visual target. It will be interesting to test in the future how stimulating climbing fibers during visual target onset affects reaching behaviors. This will reveal how a system that is responsible for providing error signals can regulate behavior online.

Despite the similarities in proving prediction error signals, inferior olive and dopamine neurons have quite different anatomic organizations in communicating with other nuclei. It seems that the cerebellum has a functionally modular structure. Each module consists of climbing fibers originated from a few inferior olive neurons that project to a specific zone of Purkinje cells, and those inferior olive neurons also receive feedback connection from the output region to which it sends projections (De Zeeuw et al., 1998; Groenewegen and Voogd, 1977; Groenewegen et al., 1979). This modular structure allows the climbing fiber to send error signals with respect to a specific movement to relevant Purkinje cells. On the other hand, dopamine neurons projecting to different areas receive largely similar inputs (Menegas et al., 2015), and single dopamine neurons have widely distributed innervation in the target (Matsuda et al., 2009). This anatomical connection makes dopamine neurons more suitable to broadcast a uniform value signal to downstream targets (Eshel et al., 2016).

The modular structure of cerebellar circuits could provide some unique advantages in studying the circuits. With modern genetic labeling strategy and virus tracing methods, as well as the development in recording techniques (e.g., deep brain calcium imaging techniques), one might be able to record the handful of neurons that belong to the same functional module and see how signals are combined and transformed at each synapse more completely.

**Advantages of mixed representation**
One surprising finding in Chapter 2 is that, in contrast to what many models predicted, few of the inputs of dopamine neurons encode pure reward or pure reward expectation signals. Instead, the activities of most neurons are modulated by both reward and reward expectations, or showed mixed coding of task-relevant variables. This type of mixed coding has been observed in many other studies and the importance of mixed selectivity has only been recognized recently (Fusi et al., 2016). The complexity of individual neuron’s response has been explained in the framework of a high-dimensional dynamical process at the level of neural population (Mante et al., 2013; Rigotti et al., 2013). This high-dimensionality representation gives the circuit more computational power to solve complex tasks (Fusi et al., 2016).

Here I provide some speculations on why mixed coding might be important for PRE computation. First, mixing of reward and expectation signals might reflect intermediate stages of RPE computation. To perform a subtractive computation, the reward and expectation signals have to be properly scaled to the same level. Mixing reward and reward expectation signals in each input area might allow finer tuning of the scales of these two signals. Second, reward representations are highly context dependent. Distributed representation of reward related signals might allow for tuning the gain of specific modalities of reward based on physiological conditions (e.g., increase the gain of the food reward and reward expectation when hungry) without interfering with the representations of other types of rewards, resulting in a more flexible control of behavior. Lastly, since animals never encounter exactly the same environment twice, the brain must learn which specific features of the sensory inputs predict the rewards. Computational algorithms that implemented reinforcement learning using a deep neural network could learn value functions accurately even though the sensory inputs are of high dimensions (Mnih et al., 2015). In our data, we saw neurons with complicated response dynamics, which
might occur in a deep reinforcement learning network. We speculate that pulling mixed signals from multiple input areas might contribute to the robustness of value representation in the face of various noise or irrelevant backgrounds. More quantitative simulations are required to demonstrate this in the future. In sum, these ideas should be explored more rigorously in biologically constrained reinforcement learning models to test for the situations in which distributed and mixed coding have advantages over encoding pure values in defined populations of neurons.

**Future directions to understand reward prediction error circuits**

As demonstrated in this dissertation, modern molecular tools such as optogenetics and virus tracing technology allow us to study neural circuits in unprecedented detail. On the other hand, with those powerful tools we must think harder than ever about what we mean by understanding a neural circuit. In his classical book (Marr et al., 2010), David Marr proposed three levels of understanding a computational system (e.g., neural circuits):

1) Computational level: what does the system try to solve?
2) Algorithmic/representational level: what algorithms and representations does the system use?
3) Implementation level: how is the system physically realized?

To conceptually map our knowledge about reward circuitry regarding these three levels: on the computational level, the reward circuitry must maximize reward and minimize punishments; on the algorithmic level, the temporal difference model can be used to learn the correct value representations; on the implementation level, dopamine neurons encode the reward prediction error signals and various downstream targets of dopamine neurons (e.g., the striatum) learn the value representation.
There are still many unanswered questions in each of these three levels. Although the computational goal of the system is to maximize rewards, we do not know how various types of stimuli from the environment are transformed into a single value that signals the amplitude of reward. In addition, given the same expected reward, human beings tend to choose certain/safe reward over risky reward (Kahneman and Tversky, 1979). This suggests that maximizing expected value of reward is not the only goal of human reward-seeking behavior. Understanding what makes safe reward more rewarding than risky reward will elucidate how reward is processed in the brain. On the level of an algorithm, the temporal difference model is a very attractive idea, although there is little direct evidence to show whether or not the brain uses this algorithm. On the other hand, behavioral studies suggest that there might be multiple learning algorithms running in parallel to generate flexible learning behaviors (Gläscher et al., 2010; Lee et al., 2012). It is possible a more elaborated version of the TD model can explain the complexity of the learning behavior, or perhaps some new types of models are required. Lastly, on the implementation level, it will be interesting to study how much the redundant information observed in each input area is due to recurrent connection. To answer this question, one has to inactivate one input area and see how the signals in other input areas are altered.

Our approach of studying prediction error computation in dopamine neurons can be extended to study other prediction error computation circuits. By comparing different circuits underlying prediction error computation, we can examine whether there are any general strategies that our brain takes to perform error correction.
REFERENCES


