Modulation of Dopamine for Adaptive Learning: a Neurocomputational Model

Jeffrey B. Inglis, Vivian V. Valentin & F. Gregory Ashby

Computational Brain & Behavior

ISSN 2522-0861

Comput Brain Behav
DOI 10.1007/s42113-020-00083-x

ONLINE FIRST

Springer

Computational Brain & Behavior
https://doi.org/10.1007/s42113-020-00083-x

ORIGINAL PAPER

Modulation of Dopamine for Adaptive Learning: a Neurocomputational Model

Jeffrey B. Inglis[1] · Vivian V. Valentin[2] · F. Gregory Ashby[2]

Abstract

There have been many proposals that learning rates in the brain are adaptive, in the sense that they increase or decrease depending on environmental conditions. The majority of these models are abstract and make no attempt to describe the neural circuitry that implements the proposed computations. This article describes a biologically detailed computational model that overcomes this shortcoming. Specifically, we propose a neural circuit that implements adaptive learning rates by modulating the gain on the dopamine response to reward prediction errors, and we model activity within this circuit at the level of spiking neurons. The model generates a dopamine signal that depends on the size of the tonically active dopamine neuron population and the phasic spike rate. The model was tested successfully against results from two single-neuron recording studies and a fast-scan cyclic voltammetry study. We conclude by discussing the general applicability of the model to dopamine-mediated tasks that transcend the experimental phenomena it was initially designed to address.

Keywords Dopamine · Adaptive learning rate · Computational cognitive neuroscience · Ventral subiculum

Introduction

Normative and machine learning models of learning have been integral to development and progress in a wide range of fields, including computer science (Sutton and Barto 1998), neuroscience (Maia 2009; Dayan and Abbott 2001), and psychology (Rescorla and Wagner 1972; Bush and Mosteller 1951; Berridge 2000). For example, reinforcement learning algorithms have provided successful models of how predicted reward estimates are updated when new rewards are encountered in the environment. In these models, the amount of learning on each trial is proportional to the reward prediction error (RPE), which is defined as the obtained reward (R) minus the predicted reward (P).

The standard assumption is that dopamine (DA) neurons in the ventral tegmental area (VTA) and the substantia nigra pars compacta (SNpc) encode the RPE via their response to rewarding events and to cues that predict rewards (Montague et al. 1996; Schultz et al. 1997). Even so, it is also well known that RPE is an imperfect predictor of the DA response. For example, DA neurons also respond to novel events and to salient stimuli with no reward-related associations (Horvitz 2002). In addition, there are large individual differences in the DA response to any given RPE, which depend, at least in part, on personality type (Pickering and Pesola 2014).

To account for such variability in the DA response to RPE, reinforcement learning models typically include an additional learning rate parameter—denoted by λ_n in Eq. 1 below—that controls the amount of learning that occurs for any given value of RPE. When fitting reinforcement learning models to data, λ_n is typically treated as a free parameter, which allows the models to account for unexplained variability in the learning effects of any given RPE, albeit only via post hoc curve fitting. A complete theory of learning must describe a neural account of these changes in λ_n. This article takes a significant step towards this goal by describing a neural network that modulates the

✉ F. Gregory Ashby
fgashby@ucsb.edu

Jeffrey B. Inglis
jeffrey_inglis@ucsb.edu

Vivian V. Valentin
valentin@ucsb.edu

[1] Interdepartmental Graduate Program in Dynamical Neuroscience, University of California, Santa Barbara, CA 93106, USA

[2] Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA 93106, USA

Published online: 12 June 2020

DA response to RPE under a wide variety of environmental conditions.

If the learning rate λ_n is too small, learning is slower than necessary and the learner is insensitive to changes in the reward structure of the environment. If λ_n is too large, learning is unstable. The optimal value of λ_n changes adaptively in response to environmental changes in the statistical structure of rewards (Daw and O'Doherty 2014; Dayan et al. 2000; Dayan and Long 1998). Additionally, a number of investigators have proposed a variety of factors that may affect λ_n such as expected and unexpected uncertainty (Dayan and Yu 2003; Yu and Dayan 2005), volatility (Behrens et al. 2007), outcome, informational, and environmental uncertainty (Mathys et al. 2011), covariance between predictions and past RPEs, estimation, and unexpected uncertainty (Payzan-LeNestour and Bossaerts 2011; Preuschoff and Bossaerts 2007), and state-feedback contingency (Crossley et al. 2013) (for detailed reviews of some of these taxonomies of uncertainty and the relationships between them, see Bland and Schaefer 2012 and Soltani and Izquierdo 2019).

In the language of Marr (1982), almost all of these models are computational. Thus, they make little or no attempt to describe the neural circuitry that implements the proposed computations. In particular, there are few current hypotheses about the neural mechanisms that modulate the amount of learning that occurs for any given RPE (for exceptions, see Bernacchia et al. 2011; Franklin and Frank 2015; Iigaya 2016; Farashahi et al. 2017).

This article proposes such a mechanism. Specifically, we describe a biologically detailed computational model of how the adaptive learning rate proposed in the models described above could be implemented at the neural level. We describe the neural circuit that mediates this modulation and model activity at the level of spiking neurons. The input to the network is a computed value of some relevant theoretical variable such as unexpected uncertainty, volatility, or feedback contingency and the output is spiking activity in a population of DA neurons. The resulting DA release is presumed to then affect tonic and phasic DA levels in target brain regions. The model is agnostic about which factors modulate learning rates and how they are computed. The neuroanatomy of the network we propose is consistent with many of the alternative proposals about how learning rates are modulated. Thus, the proposed model should be of widespread interest.

Furthermore, the model can be applied to a variety of DA-mediated tasks, many of which transcend the experimental phenomena that it was created to address. Potential applications of the model extend to working memory, creative problem solving, cognitive flexibility, and category learning. The general applicability of the model to paradigms that extend beyond the implementation of learning rates in simple reinforcement learning tasks follows from the fact that the network predicts changes in tonic and phasic DA in all brain regions that are targets of VTA DA neurons, and thus is applicable to any model of behavior that depends on these regions and assigns a specific functional role to DA.

The article begins with a brief review of a simple and common reinforcement learning algorithm. We then discuss the benefits of an adaptive learning rate and briefly review many of the factors that have been proposed that influence this rate. We refer to these factors as modulating variables. Next, we describe our neurocomputational model of how the modulating variable controls DA neuron firing and therefore also DA release and learning. The computational principle implemented by the network is to control the gain on the DA response to any given RPE in addition to regulating tonic levels via the modulating variable. This new theory is formulated as a biologically detailed computational model that we refer to as the Modulation of Dopamine for Adaptive Learning (MODAL) model. Finally, we close with a discussion of the relationship between our implementational-level model and other levels of analysis and possible directions for future research.

Reinforcement Learning Algorithms

This article proposes a neural interpretation of learning rates. Virtually all learning algorithms include a learning rate parameter and the network described below could provide a neural interpretation of that parameter in many of these algorithms. To keep the presentation concrete however, we focus on one simple reinforcement learning algorithm that is ubiquitous in the literature and that formalizes the notion of a learning rate—namely, the single-operator model of Bush and Mosteller (1951) (also see Rescorla and Wagner 1972).

The single-operator model assumes that the predicted reward value on trial n, denoted by P_n, equals:

$$\begin{aligned} P_n &= P_{n-1} + \lambda_n (R_n - P_{n-1}) \\ &= P_{n-1} + \lambda_n RPE_n, \end{aligned} \tag{1}$$

where R_n is value of the obtained reward on trial n and λ_n is the learning rate on trial n. It is well known that in a stable environment, P_n converges asymptotically to the mean reward value and the rate of convergence increases with λ_n (i.e., for all λ_n in the range $0 < \lambda_n < 1$). So any variable that increases λ_n, increases the learning rate.

Note that even simple algorithms that set all λ_n to the same constant value predict a form of cooling because the magnitude of the RPEs will decrease as learning progresses. Even so, many algorithms change λ_n with n (e.g., Sutton 1992). For example, it is common to decrease

λ_n as n increases—a process that accelerates cooling. In addition, there have been many proposals that other factors also dynamically adjust learning rates in the brain. The remainder of this section briefly reviews various modulating variables that have been proposed to affect λ_n.

Although λ_n is often treated as a free parameter in many applications, its optimal value can be determined trial-by-trial by considering the iterative updates in reinforcement learning as a statistical problem of how best to integrate previous estimates with new evidence. Taking a Bayesian approach, it has been shown that under certain assumptions, the optimal way to integrate past predictions with new data is to set the learning rate to (Daw and O'Doherty 2014; Dayan and Long 1998; Dayan et al. 2000):

$$\lambda_n = \frac{\sigma_n}{\sigma_n + Var(R)}, \tag{2}$$

where σ_n represents the variance or uncertainty in our current estimate of the predicted reward and $Var(R)$ represents the variance in the reward values. Therefore, if the obtained reward values are not changing very much (i.e., $Var(R)$ is small), then λ_n should be large, which will cause the predicted reward estimate to converge quickly to the (mean) obtained reward value. On the other hand, if the obtained reward values are noisy (i.e., $Var(R)$ is large), then we should set λ_n to be small to avoid over-reacting to an unexpectedly large or small reward value.

Several researchers have argued that learning rates in the brain are also affected by volatility—that is, by how quickly the reward contingencies change in the environment (Behrens et al. 2007; Mathys et al. 2011). The idea is that increases in volatility should increase λ_n because agents should learn faster in a rapidly changing environment in order to track the fluctuations. Alternatively, when the environment is stable, the agent should learn more slowly to ensure it uses as much data as possible in order to converge upon the true stable reward probabilities.

Dayan and Yu proposed that learning rates depend on what they called expected and unexpected uncertainty (Yu and Dayan 2005; Dayan and Yu 2003). Expected uncertainty arises as a result of the unreliability of the cue that signals reward and the agent should suppress the use of the cue when expected uncertainty is high. Unexpected uncertainty is similar to the Behrens et al. (2007) notion of volatility and the Mathys et al. (2011) notion of environmental uncertainty, that is, unexpected uncertainty is high when the agent is confident in their top-down model but these expectations are nonetheless violated by the bottom-up sensory data. This may be an indication that although the model was accurate, the environment has changed and therefore learning from bottom-up data should be more heavily weighted than the top-down model. However, according to Bland and Schaefer (2012), unexpected uncertainty differs from volatility in that volatility is related to the frequency with which stimulus-response-outcome (SRO) contingencies change. For example, in a probabilistic reversal task where SRO contingencies reverse every 30 trials, unexpected uncertainty will increase following the reversal. Furthermore, this environment would be characterized as having higher volatility relative to an environment in which the SRO contingencies only reversed every 100 trials.

Payzan-LeNestour and Bossaerts (2011) proposed that λ_n depends on unexpected uncertainty and estimation uncertainty and that prediction risk scales the RPE (Preuschoff and Bossaerts 2007). Prediction risk is the irreducible uncertainty due to outcome uncertainty. Estimation uncertainty is measured as the entropy of the posterior distribution (similar to the uncertainty of the prior in the above equation), whereas unexpected uncertainty is high when SRO contingencies change abruptly, as described above. Preuschoff and Bossaerts (2007) also proposed that the covariance between past predictions and reward prediction errors may contribute to λ_n, as derived from least-squares learning theory.

Finally, empirical evidence suggests that state-feedback contingency, defined as the covariance between rewards and predictions, has a significant effect on the learning rate (Ashby and Vucovich 2016; Crossley et al. 2013). The intuition here is that a measure of the covariance between rewards and predictions enables a parsimonious method for the agent to infer the degree to which its actions play a role in determining its rewards. If state-feedback contingency is high, the agent recognizes that its behavior plays a significant role in determining its rewards and takes advantage of this by increasing the learning rate. Alternatively, if state-feedback contingency is low, the agent recognizes that its behavior does not play a significant role in determining its rewards and therefore it can conserve resources and preserve previous learning by decreasing the learning rate.

The next section proposes a neural network that could implement any of these modulating effects on learning rate.

Neuroanatomy of MODAL

Reward and feedback processing recruit diverse brain networks that include the limbic system and prefrontal and sensory cortices (Liu et al. 2011; Watabe-Uchida et al. 2012; Tian and Uchida 2015; Haber 2016; Faget et al. 2016; Takahashi et al. 2016). Multiple brain regions respond to reward and compute predicted rewards (Sesack and Grace 2010; Bromberg-Martin et al. 2010; Humphries and Prescott 2010), and this redundancy inspired many alternative theories of how DA neuron firing is modulated by RPE (Houk et al. 1995; Schultz et al. 1997; Sutton and Barto 1998; Schultz 1998; Tan and Bullock 2008;

Kawato and Samejima 2007; Morita et al. 2013; Brown et al. 1999; Hazy and Frank 2010; Joel et al. 2002; Stuber et al. 2008; Contreras-Vidal and Schultz 1999; O'Reilly et al. 2007). In contrast to all this work, we do not know of any neurocomputational models that attempt to account for any modulating effects of the DA response to RPE. We propose that dynamic changes in learning rate are mediated by changes in the size of the population of tonically firing DA neurons. As the size of this population grows, more DA neurons become available to respond to any given RPE, which has the computational effect of increasing the learning rate.

The neural architecture of the model is described in Fig. 1. The inputs to the network are from regions that compute RPE and the value of the relevant modulating variable. Whereas the alternative modulating variables that have been proposed might recruit somewhat different neural networks, they all depend on temporal integration or continuous updating of feedback and reward information. Therefore, they are likely to depend on similar networks that include regions in orbitofrontal, medial prefrontal, anterior cingulate, parahippocampal, and entorhinal cortices. We make no attempt to describe this network in detail, but we assume that it sends a prominent projection to the ventral subiculum (vSub), which is the main output structure of the hippocampus. vSub receives input from a variety of regions, including CA1 of the hippocampus (Fanselow and Dong 2010), parahippocampal cortex, and entorhinal cortex (Kerr et al. 2007). The entorhinal cortex encodes general properties of the current context (Jacobs et al. 2010), and the parahippocampal cortex has a general role in contextual binding (Aminoff et al. 2013). Additionally, the entorhinal cortex receives almost all of its cortical inputs from polymodal association areas, including cingulate, orbitofrontal, and parahippocampal cortices, making it well situated for integrating diverse inputs (Insausti et al. 1987). Given the positioning of the vSub as an interface between the contextual information processing in the hippocampus and cortical and subcortical regions implicated in reward processing, learning, and motivation (Quintero et al. 2011), we propose that the vSub is a likely target of the complex neural networks that mediate processing of the alternative modulating variables that have been proposed.

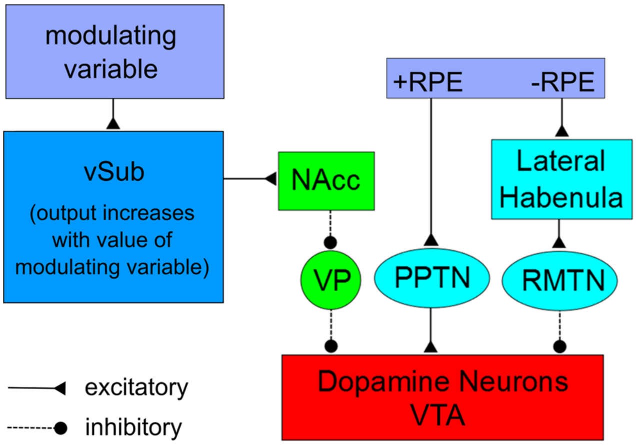

Fig. 1 Neural architecture of the proposed MODAL model of how DA activity is regulated by one of the modulating variables described in the previous section. RPE, reward prediction error; vSub, ventral subiculum; NAcc, nucleus accumbens; VP, ventral pallidum; PPTN, pedunculopontine tegmental nucleus; RMTN, rostromedial tegmental nucleus; VTA, ventral tegmental nucleus

The right half of the Fig. 1 network instantiates the standard RPE model. The idea is that reward sensitive units in regions such as prefrontal and orbitofrontal cortex contribute to the RPE DA signal by providing excitatory inputs to the pedunculopontine tegmental nucleus (PPTN) (Hong and Hikosaka 2014; Kobayashi and Okada 2007; Okada and Kobayashi 2013) and lateral habenula (LH) (Tian and Uchida 2015; Hong et al. 2011; Matsumoto and Hikosaka 2007, 2009). Through these circuits, positive RPEs excite VTA DA neurons via the PPTN, whereas negative RPEs inhibit VTA DA activity via the LH (and the rostromedial tegmental nucleus (RMTN)).

The more novel features of the Fig. 1 model are presented in the left half of the figure. First, factors thought to influence the modulating variable are integrated in the vSub, which results in an output signal to the NAcc that is proportional to the value of the modulating variable.

The next component of the model builds on the work of Grace et al. (2007), who proposed that the pathway vSub → NAcc → VP → VTA controls the number of VTA DA neurons that fire tonically. The NAcc → VP and VP → VTA projections are both GABAergic, but the tonic firing rate of VP neurons is much higher than the tonic firing rate of NAcc neurons. As a result, many DA neurons in VTA are silent due to tonic inhibition by VP. Estimates suggest that because of this inhibition, only about half of VTA DA neurons are spontaneously active under control conditions, and these tonically firing neurons are the only ones available for phasic bursts when excited by PPTN (Lodge and Grace 2006). When the value of the modulating variable is high, vSub excites NAcc neurons, which inhibit VP neurons. This releases VTA DA neurons from tonic inhibition, which increases the number of tonically firing VTA DA neurons, thereby enlarging the pool of DA neurons that can respond to excitatory input from PPTN. In this way, increasing the value of the modulating variable amplifies the RPE-induced VTA DA response. Thus, the Fig. 1 network proposes a neural mechanism via which the modulating variable can control the gain of the DA response to any given RPE.

To test this theory more rigorously, we built a biologically detailed computational model of the Fig. 1 network, and examined its ability to account for RPE and learning rate effects on DA release. Our model is consistent with known neuroanatomy and neurophysiology and accurately accounts for neuroscientific data.

Neurocomputational Details

We built a computational cognitive neuroscience model of the Fig. 1 network that includes spiking neurons as the basic units and that obeys the relevant neuroscience constraints (e.g., Ashby 2018; Ashby and Helie 2011). The model was programmed using the Python programming language (Van Rossum and Drake 2011).

Model Architecture and Activation Equations

As described earlier, a rough schematic of MODAL is shown in Fig. 1. Our main goal was to understand how changes in the value of the modulating variable affect the rate of learning via modulations of VTA DA neuron activity. As a result, we made no attempt to model neural firing in hippocampus or upstream cortical regions. Modeling these complex structures is beyond the scope of the current project. Furthermore, we modeled activation in vSub, PPTN, and LH as either on or off (via square waves). Because our hypothesis is that the value of the modulating variable affects VTA DA neuron activity via the NAcc → VP → VTA pathway, we modeled activity in all these structures using spiking-neuron models; specifically Izhikevich (2007) medium spiny neuron (MSN) models for NAcc, quadratic integrate-and-fire models for VP (Ermentrout 1996), and Izhikevich (2003) regular spiking neuron models for VTA. Parameter values for the Izhikevich units were set equal to the values used by Izhikevich (2007) and parameter values for the quadratic integrate-and-fire units were identical to those used in Ashby (2018), except when otherwise noted.[1]

Postsynaptic effects of a spike were modeled via the α-function (e.g., Rall 1967). Specifically, when the presynaptic unit spikes, the input projected to the postsynaptic unit is (with spiking time $t = 0$):

$$\alpha(t) = \frac{t}{\delta} \exp\left(1 - \frac{t}{\delta}\right). \tag{3}$$

The parameter δ, which models temporal delays in synaptic transmission, was set to 123 for NAcc and VP units, and 225 for VTA units.

The following subsections describe additional details about how we modeled activity in NAcc, VP, and VTA. Table 1 lists values of all connectivity parameters. These parameter values were based on biological constraints (e.g., excitatory versus inhibitory). In its current form, MODAL does not exhibit any synaptic plasticity; therefore, all connection weight parameters in this network were fixed throughout the simulations.

NAcc

The NAcc layer was modeled with 100 Izhikevich (2007) MSNs with input to NAcc$_i$ (for $i = 1, 2, ..., 100$):

$$I_{NAcc_i}(t) = vSub(t), \tag{4}$$

where $vSub(t)$ represents activation in vSub as a square wave with amplitude equal to the value of the modulating variable. For simplicity, the tonic firing rate of NAcc in the absence of input was chosen to be 0 Hz, which is reasonable considering that Fabbricatore et al. (2009) reported a tonic rate of 0.53 Hz.

Braganza and Beck (2018) hypothesized that the disinhibition motif that characterizes the basal ganglia plays the computational role of gating. However, in addition to gating DA via disinhibition, MODAL does this in a continuous fashion such that as the value of the modulating variable increases, the size of the population of VTA DA neurons also increases. In other words, whereas the disinhibition motif implements gating at the single synapse level, at the population level it can implement a gain or amplification of the signal. The striatal MSNs provide an excellent candidate for implementing the amplification. The MSNs exhibit bistable dynamics consisting of up and down states. In the up state these neurons are responsive to inputs and will fire spikes, while in down states they tend not to fire in response to inputs.

In MODAL, the NAcc neurons play a critical role in controlling the size of the population of tonically active VTA DA neurons (i.e., because of their one-to-one connectivity through the VP). When a NAcc neuron transitions from its down state to its up state, the VP neuron it projects to is silenced due to NAcc inhibition. This releases the corresponding VTA neuron from tonic inhibition, causing it to fire tonically and become responsive to inputs from PPTN and LH. Therefore, although the NAcc neurons are also responsive to inputs while in their up state,

Table 1 Connectivity parameter values between layers of MODAL

Parameter	Value
$w_{PPTN \rightarrow VTA_i}$	125
$w_{LH \rightarrow VTA_i}$	−125
$w_{NAcc_i \rightarrow VP_i}$	−10
$w_{VP_i \rightarrow VTA_i}$	−1000

[1] However, note that in Izhikevich (2007) and Ashby (2018), the β parameter controls the rate of tonic spiking. Each region in our model has a different tonic firing rate; therefore, $\beta = 0$ in NAcc, $\beta = 20$ in VP, and $\beta = 62$ in VTA.

their key role in MODAL is to determine the appropriate size of the active VTA DA neuron population.

A key property of MODAL is that the size of the population of tonically firing DA neurons grows with increases in the value of the modulating variable. This requires NAcc neurons to transition from their down states to their up states at different levels of input to the vSub. To implement this property, each NAcc neuron in the model has a different resting state drawn randomly from a uniform distribution between − 93.5 and − 55 mV. Figure 2 shows the effect of the different resting states on the nullclines and state trajectories of three NAcc neurons (for more detail on the application of dynamical systems theory and phase portraits to neural modeling, interested readers may consult Izhikevich 2007 or Ashby 2018). The top row of Fig. 2 shows predicted intracellular voltage levels for three NAcc neurons and the bottom row shows the corresponding phase portraits. The neurons are all identical, except for their resting membrane potential, which is low in column 1 (− 93.5 mV), medium in column 2 (− 75 mV), and high in column 3 (− 55 mV). Notice that increasing the level of vSub activation causes all the v-nullclines to shift upwards. For the neuron with the lowest resting state (− 93.5 mV, left), this upward shift is not sufficient to cause the neuron to undergo a saddle-node bifurcation and therefore the state moves from the fixed point 1 to point 2 on the v-nullcline and it slides down the v-nullcline until it reaches the new fixed point (3) and the down state persists due to insufficient input from vSub (the neuron does not spike). Alternatively, for the neuron with the intermediate resting state (− 75 mV, center), the upward shift in the v-nullcline is sufficient to cause the neuron to undergo a saddle-node bifurcation, moving the state from the fixed point 1 to point 2 on the v-nullcline. Due to the ghost of the saddle-node, the state slides slowly along the v-nullcline until point 3 when it leaves the v-nullcline and the derivative of v goes positive causing the voltage to increase rapidly, leading to a transition to the up state and spiking behavior. Once a spike is registered (4), the voltage is reset (5) below the ghost of the saddle-node leading to shorter latency spikes (6). The neuron with the highest resting state undergoes similar behavior to the intermediate neuron, except the latency to the spike is substantially shorter. This is because the upward shift of the v-nullcline results in point 2 being below the v-nullcline, immediately leading to a positive derivative of v and rapid transition to the up state and spiking (3). Furthermore, following a spike, the voltage is reset below the v-nullcline (and the ghost) (4) and therefore subsequent spikes are rapid (5).

The key result of this network architecture is that there is a continuum where neurons with lower resting states require substantially more current to undergo the saddle-node

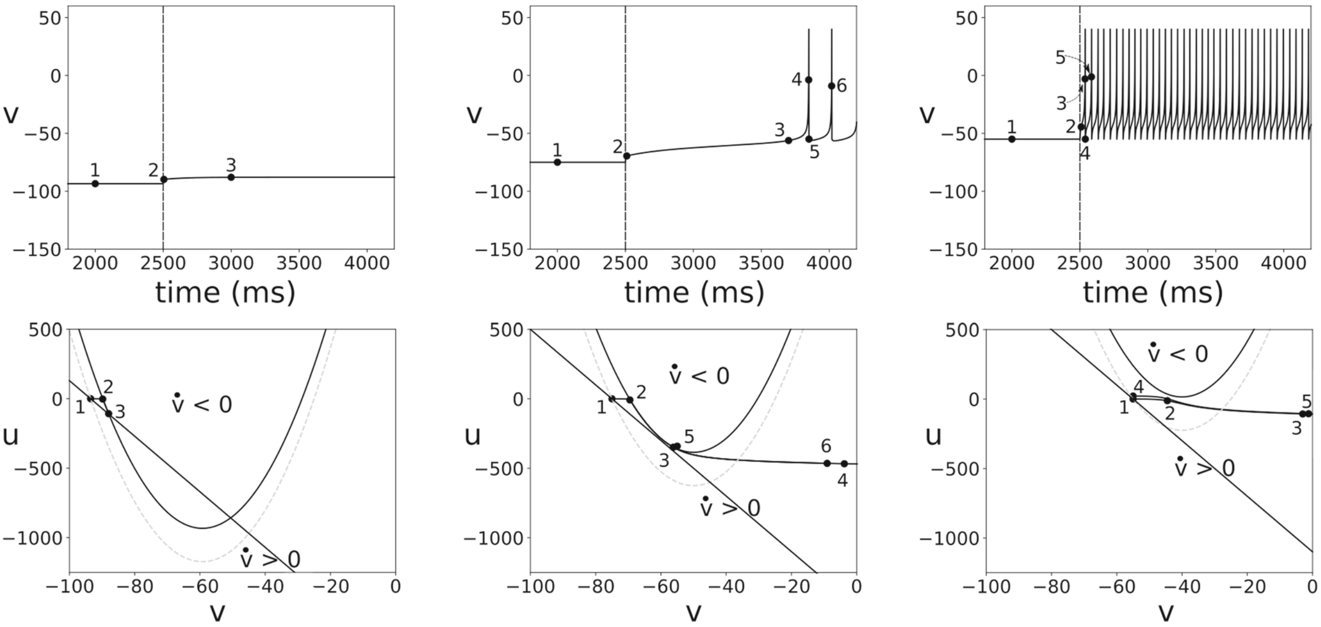

Fig. 2 Top: spiking behavior of three NAcc medium spiny neurons with low (− 93.5 mV, left), intermediate (− 75 mV, center), and high (− 55 mV, right) resting potentials. The vertical dashed line represents the time when the current is injected into the neuron. Bottom: Phase portraits with v and u nullclines and state trajectories for each of the neurons in the top panel. The v-nullcline is represented by the gray dashed line (no current) and the black u-shaped curve (current on). The u-nullcline is represented by the straight black line. When the current is turned on the v-nullcline shifts upwards. The numbers in the bottom plots correspond to the time points indicated by the numbers in the plots directly above

bifurcation, relative to neurons with higher resting states. This has the desired effect of increasing the size of the VTA DA population as the value of the modulating variable increases by transitioning more NAcc neurons into their up states.

VP

The VP layer was modeled with 100 quadratic integrate-and-fire units. The input to VP$_i$ ($i = 1, 2, ..., 100$) was equal to:

$$I_{VP_i}(t) = w_{NAcc_i \to VP_i} \times \alpha_{NAcc_i}(t) \tag{5}$$

where $w_{NAcc_i \to VP_i}$ is the connection weight between NAcc$_i$ and VP$_i$ and $\alpha_{NAcc_i}(t)$ is the integrated α-function generated by spikes in NAcc$_i$.

The tonic firing rate for VP units was set to approximately 7 Hz, which is consistent with measurements reported by Root et al. (2012). Despite NAcc inhibition, the higher tonic firing rate of the VP units relative to the NAcc units has the effect of ensuring that the VP units still fire spikes at low values of the modulating variable. As the modulating variable increases, more NAcc neurons transition to their up states, silencing more VP units. Each VP unit is connected to one VTA unit.

VTA

The DA neurons in the VTA were modeled with 100 Izhekivich regular-spiking neurons. Call these units VTA$_i$ ($i = 1, 2, ..., 100$). All 100 units received identical input from PPTN and LH. In addition, the VTA$_i$ unit received input from the corresponding VP$_i$ unit.

The input to unit VTA$_i$ was:

$$I_{VTA_i}(t) = \begin{cases} w_{VP_i \to VTA_i} \alpha_{VP_i}(t) \\ + w_{PPTN \to VTA} PPTN(t) + w_{LH \to VTA} LH(t) \end{cases} \tag{6}$$

where $w_{VP_i \to VTA_i}$ denotes the synaptic strength between VP$_i$ and VTA$_i$, $\alpha_{VP_i}(t)$ denotes the output of unit VP$_i$ at time t (i.e., the α-function), $w_{PPTN \to VTA}$ denotes the synaptic strength between PPTN and all VTA neurons and $w_{LH \to VTA}$ denotes the synaptic strength between LH and all VTA neurons.

Activation in PPTN was modeled as follows:

$$PPTN(t) = \begin{cases} RPE & \text{if } RPE > 0 \text{ and } 7000 \le t < 7100 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

This results in a square wave with amplitude equal to RPE (for positive RPE) lasting 100 ms (Bayer et al. 2007).

Activation in LH was:

$$LH(t) = \begin{cases} 1 & \text{if } RPE < 0 \text{ and } 7000 \le t < (7000 - 400RPE) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

This results in a square wave of amplitude equal to 1 (for negative RPE) of varying duration (0 – 400 ms) that elicits pauses in VTA units with the length of the pause proportional to the magnitude of the negative RPE.[2] This formulation of activity in PPTN and LH produces DA neuron firing that is proportional to RPE (Bayer and Glimcher 2005) and results in symmetric encoding of positive and negative RPE by extracellular DA concentrations (Hart et al. 2014).

Single-Neuron Dynamics of MODAL

The dynamics of three neurons in each layer of MODAL are illustrated in Fig. 3. The level of input to vSub increases linearly from 0.0001 to 1 in increments of 0.0001. At the beginning of the 10-second interval, vSub activation is low and all of the NAcc neurons are in their down state. However, as vSub input increases slightly, the first NAcc unit from the left transitions to its up state and increases its firing rate, which silences the first VP unit. This disinhibits the first VTA unit and it begins to fire tonically, making it possible for this unit to burst or pause in response to input from PPTN or LH. Similarly, as vSub input increases further, the second (center) and then the third (right) NAcc units also transition to their up states, which first silences the second and then the third VP units, respectively. This causes the second and then the third VTA units to become disinhibited and fire tonically, making them available for bursting or pausing as well. This network structure creates the desired effect of having a larger pool of VTA DA units available for bursting and pausing as the level of vSub input rises. The level of vSub activation that determines the size of the VTA DA neuron population depends on the preferred resting states of each of the NAcc neurons. The PPTN activation in this simulation alternates between 0 and 1 for 1000-ms intervals. Notice that as long as the corresponding NAcc neuron is silent, the VTA neuron is unresponsive to inputs from PPTN. However, once the NAcc firing rate is

[2] However, for Figs. 3, 4, 5 (left and center), and 6, the PPTN square wave lasted 1000 ms and the LH square wave lasted a maximum of 1000 ms. This was done to ensure a sufficiently long interval to extract accurate measurements of firing rate and active population size. Figures showing dopamine output used the parameters described in the text.

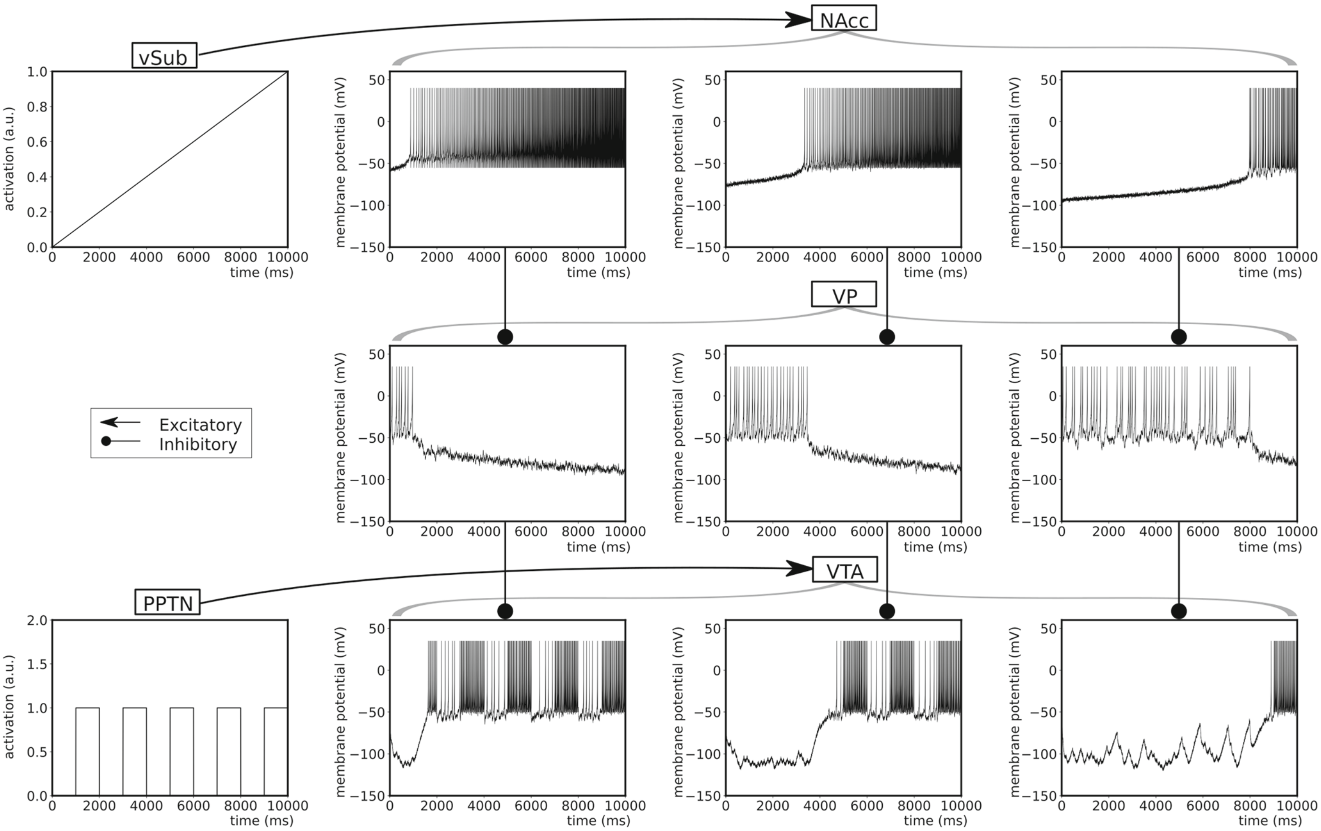

Fig. 3 Spike dynamics in the simplified network for a 10-s interval. As activation in vSub increases this causes more NAcc neurons to transition into their up state, thereby inhibiting more VP neurons and disinhibiting more VTA neurons. The result of this architecture is a recruitable pool of VTA DA neurons that can respond to inputs from PPTN and LH (not shown). See text for more details on the dynamics. vSub, ventral subiculum; NAcc, nucleus accumbens; VP, ventral pallidum; VTA, ventral tegmental area; PPTN, pendunculopontine nucleus

high enough to disinhibit VTA neurons, they now alternate between periods of tonic firing and phasic bursting (or tonic firing and phasic pausing, not shown in Fig. 3).

Methods

The proposed model was evaluated using numerical simulations on two different types of data from nonhuman animals: single-unit recordings and fast-scan cyclic voltammetry. The goal of the simulations was to test the neural architecture of the model, not parameter optimization. Hence, it is important to note that although this network includes a number of parameters, the majority of these were fixed after modeling each level of experimental data. In particular, the parameters that were estimated when fitting the single-unit data of Lodge and Grace (2006) then remained fixed at those values in all future simulations. This process ensured that the network is able to account simultaneously for experimental data at many levels of analysis and implicitly implements a significant degree of inflexibility into the model's structure by constraining it by the lower levels of analysis. The key parameters that were modified will be described in each section. For all simulations, the voltage for each unit was estimated for each millisecond of a 10,000-ms trial.

In the neural simulations, there was no learning and noise was minimal; therefore, results are from a single simulation. We ran the neural network through simulations with 100 levels of the value of the modulating variable (from 0.01 to 1 by increments of 0.01) and 201 values of RPE (from − 1 to 1 by increments of 0.01). The amount of DA released by the network was computed for each combination of learning rate and RPE, resulting in a total of 20,100 DA measurements.

Results

MODAL was subjected to three neural benchmark tests. First, we explored whether it could account for the Lodge and Grace (2006) results showing that vSub activation increases the number of tonically active VTA DA neurons, whereas activation of the PPTN induces burst firing of VTA

DA neurons. Second, we examined whether the model was consistent with the data of Bayer and Glimcher (2005), which showed that DA neuron firing increases linearly with RPE between minimal and maximal values. Third, we tested whether the model could account for the data of Hart et al. (2014), which showed that DA release (i.e., extracellular DA concentration) is a linear function of RPE and that positive and negative RPEs are encoded symmetrically.

Neural Tests of MODAL

Benchmark Test 1: Distinct Pathways in PPTN and vSub

Lodge and Grace (2006) provided evidence that distinct interacting pathways exhibit differential influences on VTA DA neurons. In this experiment, they activated vSub, PPTN, or both structures via NMDA infusion, and then they counted the number of VTA DA neurons that were firing tonically (i.e., per electrode track), and they also estimated the average firing rate of all VTA DA neurons. Their results are shown in the left column of Fig. 4. Note that vSub activation increased the number of tonically active VTA DA neurons but did not affect the population firing rate. On the other hand, activation of the PPTN caused an increase in the population firing rate as a result of burst firing, but did not affect the size of the tonically active population. Finally, simultaneous activation of vSub and PPTN caused a significant increase in both burst firing and the size of the tonically active population.

MODAL fits are shown in the right column of Fig. 4. We simulated the control condition of Lodge and Grace (2006) by setting the square-wave activations of vSub, LH, and PPTN to 0.27, − 0.31, and 0, respectively. Activation of vSub by NMDA infusion was simulated by changing the amplitude of the square wave activation of vSub to 1. Activation of PPTN was simulated by changing the amplitude of the square-wave activation of PPTN to 0.05. Note that the model accurately captures all qualitative features of the data. It should also be noted that the connection weights between PPTN and the VTA units were set so that excitatory input from PPTN to a tonically firing VTA neuron was sufficient to result in burst firing according to the criteria used by Lodge and Grace (2006) [i.e., an interspike interval (ISI) of ≤ 80 ms and bursting that persists until the ISI exceeds 160 ms; (Grace and Bunney 1983)].

Figure 5 shows heat-maps that depict the number of active neurons and population firing rate as a function of vSub activation and RPE (Fig. 5 left and center, respectively). These plots show that the overall qualitative behavior reported by Lodge and Grace (2006) is implemented in MODAL across a wide range of input values for vSub activation and RPE. The population size of tonically firing DA neurons increases with vSub activation, but is relatively independent of RPE, and therefore of PPTN/LH activation while population firing rate changes with PPTN/LH activation but is relatively independent of vSub activation. Furthermore, Fig. 5 (right) shows the predicted extracellular DA concentration for the Lodge and Grace (2006) experiment, with minimal DA release when both vSub and PPTN activation are low, limited DA release when only one of vSub or PPTN has high activation, and maximal DA release for concurrent high activity in vSub and PPTN. We assumed extracellular DA concentrations would be proportional to the total postsynaptic effects of all DA units in the model and so we estimated extracellular DA concentrations as the integral of each VTA neuron's α-function for a 1-s

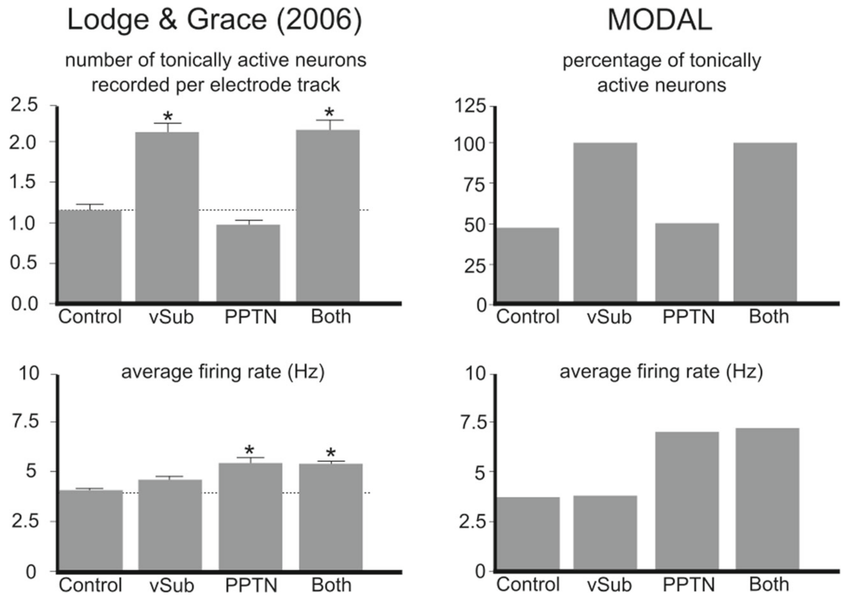

Fig. 4 Benchmark test 1. Left panel: Experimental data from Lodge and Grace (2006). Right panel: MODAL simulations of the same experiment. Plots of the experimental data are reprinted and modified from Lodge and Grace (2006). vSub, ventral subiculum; PPTN, pedunculopontine nucleus

period following reward and summed over all neurons in the population.

Benchmark Test 2: Single-Unit Recordings from DA Neurons

Bayer and Glimcher (2005) recorded from single midbrain DA neurons (VTA and SNpc) while monkeys performed a task that required them to learn to make appropriately timed eye movements. Correct responses were rewarded with a small amount of juice. Bayer and Glimcher (2005) found that the response of the midbrain DA neurons was proportional to an estimate of the RPE (the difference between obtained reward value and a weighted average of previous reward values). Their results from a population of midbrain DA neurons are shown in the left panel of Fig. 6. Note that the increase in firing rate is linear after RPE exceeds a minimum value (i.e., of around − 0.1).

Simulations of the model under similar conditions are shown in the right panel of Fig. 6. Note that the model accurately captures the qualitative properties of the data. In these simulations we set the tonic firing rate of VTA neurons to approximately 5 Hz (to match data reported by Bayer et al. 2007). For simplicity, the LH → VTA and PPTN → VTA connection weights were set to be equal, which was sufficient to cause VTA neurons to pause in response to negative RPEs. The right panel of Fig. 6 shows the average firing rate of the VTA population.

Benchmark Test 3: Extracellular DA Levels in NAcc

The DA neuron firing-rate data shown in Fig. 6 suggest a more limited dynamic range for encoding negative as opposed to positive RPEs. In particular, the amount of increase in firing rate observed for positive RPEs was considerably greater than the amount of decrease seen for negative RPEs. Bayer and Glimcher (2005) speculated that negative RPEs might also be encoded by pause duration, and Bayer et al. (2007) later reported evidence supporting this hypothesis. Of course, synaptic effects of DA are more closely related to extracellular DA levels than to DA neuron spiking. For this reason, Hart et al. (2014) used fast-scan cyclic voltammetry to examine how extracellular DA levels in the rat NAcc varied as a function of RPE. Their results are summarized in the left panel of Fig. 7. Note that the phasic bursting and pausing of midbrain DA neurons results in symmetric encoding of positive and negative RPEs in extracellular DA concentrations.

Our third benchmark test was to ask whether a model constrained by benchmark tests 1 and 2 could also account for the symmetric encoding of positive and negative RPEs shown in Fig. 7. Therefore, we simulated performance of the model in the Hart et al. (2014) experiment by choosing the maximum duration of LH activation to be 400 ms (see Eq. 8). All other parameter estimates from benchmark tests 1 and 2 were fixed. As in benchmark 1, we assumed that extracellular DA concentrations would be proportional to the total postsynaptic effects of all VTA units in the model and so we estimated extracellular DA concentrations as the integral of each VTA neuron's α-function for a 1-s period following reward and summed over all neurons in the population.

The results are shown in the right panel of Fig. 7 for a variety of different levels of vSub activation. Note that MODAL accounts for the symmetric encoding of positive and negative RPEs seen in the Hart et al. (2014) data, and it does this for all levels of vSub activation. But note that the model also makes an important novel, and to our knowledge, untested prediction – decreasing the level of vSub activation (via decreases in the value of the modulating variable) should decrease the slope of the regression line that best fits the observed extracellular DA concentrations.

The Hart et al. (2014) results shown in the left panel of Fig. 7 were averaged across results from three conditioning tasks – two that used probabilistic feedback and one that used deterministic feedback. Note that for many of the proposed modulating variables, probabilistic and deterministic feedback would likely lead to predictable differences. Therefore, our model predicts that the linear relationship evident in Fig. 7 is likely a result of averaging across three distinct linear curves.

Although not evident in Fig. 7, note that MODAL also predicts that decreases in vSub activation should result

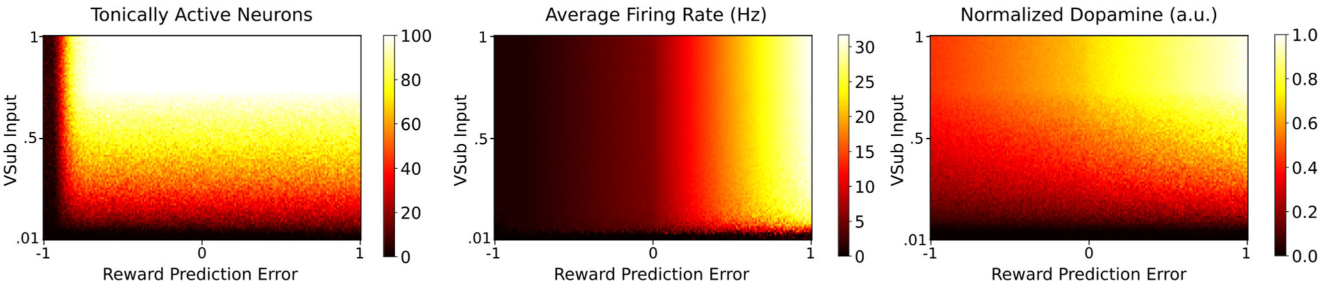

Fig. 5 Heatmaps showing the number of tonically active DA neurons (left), average population firing rate (center), and normalized DA release as a function of vSub input and RPE (right)

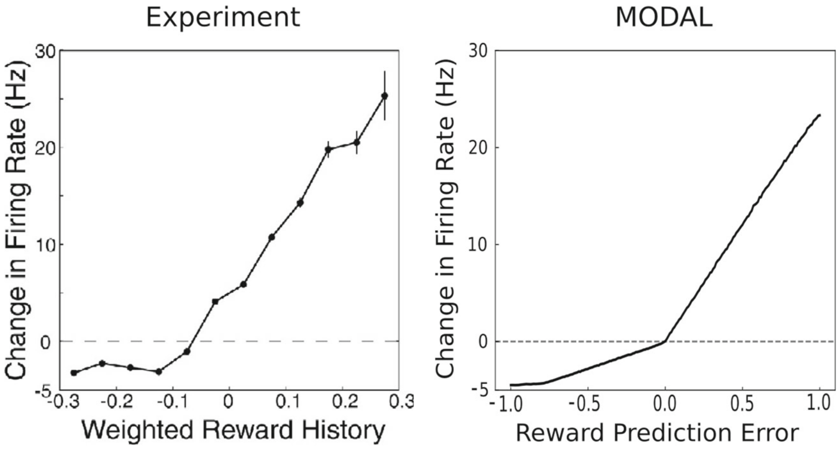

Fig. 6 Benchmark test 2. Left panel: Firing rate of a population of midbrain DA neurons as a function of RPE (from Bayer and Glimcher 2005). Right panel: MODAL simulations of the same experiment. Plots of experimental data are reprinted and modified from Bayer and Glimcher (2005)

in a downward shift in baseline or tonic concentrations of extracellular DA. This is because a reduction in vSub activation reduces the number of VTA units that are tonically firing, which reduces the number of VTA units contributing to the baseline concentration of extracellular DA.

Discussion

This article proposed a neurobiologically detailed spiking neural network model that varies the size of the population of tonically firing DA neurons in response to environmental changes. The model makes specific quantitative predictions about how changes in the size of this population alter baseline DA levels and the gain on the DA response to any given RPE. This new model successfully accounts for two single-cell recording data sets and results from a fast-scan cyclic voltammetry study.

A strong theory should make novel predictions. We highlighted two novel predictions of the model proposed here. First, any experimental change that reduces the value of the modulating variable should reduce the magnitude of change in NAcc extracellular DA concentrations for any given change in RPE. This prediction is illustrated in Fig. 7. Second, the model predicts that decreasing the value of the modulating variable should decrease tonic extracellular DA levels due to the decreased size of the active VTA DA neuron population. To our knowledge neither of these predictions of MODAL has been tested. It should be noted however that recent evidence suggests that testing this latter prediction may be complicated by effects of local mechanisms on extracellular DA concentrations (Berke 2018).

Behavioral Applications

The MODAL network illustrated in Fig. 1 includes no motor units, nor any units associated with motor planning or decision making. As a result, in its current form, MODAL produces no behavior and therefore, without some significant augmentation, it cannot be tested against behavioral data. Even so, MODAL makes strong predictions about how DA levels will vary trial-by-trial in any brain

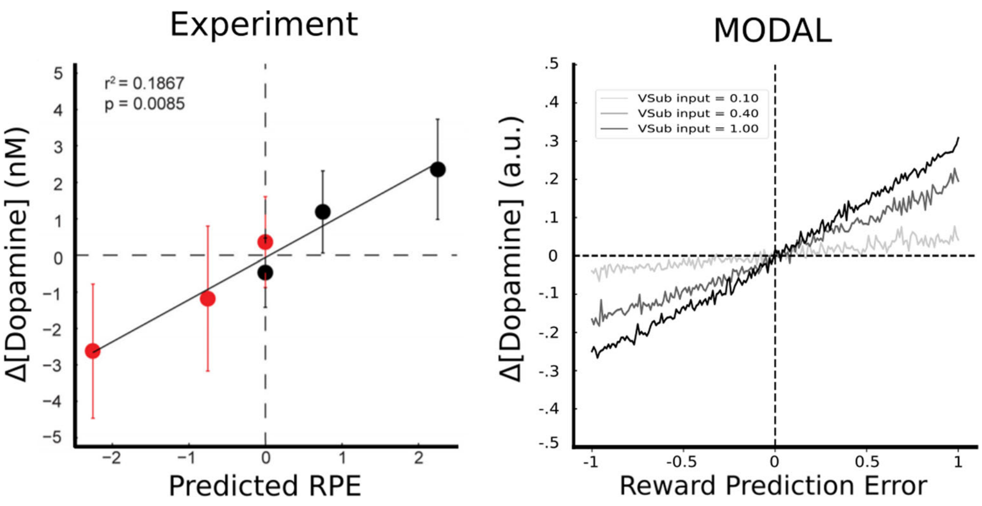

Fig. 7 Benchmark Test 3. Left panel: Experimental measurements of extracellular DA concentrations in NAcc as a function of RPE (from Hart et al. (2014)). Right panel: MODAL simulations of the (Hart et al. 2014) experiment for a variety of different levels of the modulating variable

region that is a target of VTA DA neurons. This includes regions such as prefrontal cortex, hippocampus, amygdala, ventral striatum and the most anterior portions of the dorsal striatum (e.g., head of the caudate nucleus). Therefore, MODAL could be combined with any model that accounts for behavior with a neural network that includes these regions and assigns a functional role to DA. The result should be a more powerful model of the behavior that can dynamically adjust tonic and phasic DA release in response to environmental changes in some modulating variable such as volatility, environmental uncertainty, or feedback contingency. Many such models have been proposed—far too many to review here. Furthermore, because DA projections are diffuse, rather than synapse specific, MODAL should be able to interface with a wide variety of computational models—not just those that include a high level of biological detail.

This section briefly discusses three qualitatively different types of behavioral applications of MODAL: (1) to models of value learning that could benefit from a more accurate model of reward-driven phasic DA firing; (2) to models of executive function that posit a modulatory role for cortical DA; and (3) to models of procedural learning in which synaptic plasticity depends on DA neuron activity in the substantia nigra pars compacta (SNpc).

The primary motivation for the creation of MODAL is to provide a neurocomputational mechanism for how changes in the environment can modulate the learning rate (i.e., λ_n in Eq. 1). Takahashi et al. (2008) proposed a model in which the ventral striatum encodes state values similar to those generated by Eq. 1. MODAL could be conjoined with this model since the ventral striatum is a primary target of VTA DA neurons. Furthermore, it has also been reported that the ventral striatum plays a key role in probabilistic reversal learning (Cools et al. 2002). Behrens et al. (2007) proposed a Bayesian model of reversal learning in which the learning rate changes with the volatility of the environment. Their model is purely computational and makes no attempt to describe any of the underlying neural circuitry. Therefore, MODAL could be integrated with the Behrens et al. (2007) model to produce a more biologically detailed model of reversal learning.

Within the striatum, DA is quickly cleared from synapses by DA active transporter (DAT) and, as a result, the temporal resolution of DA in the striatum is high enough so that DA levels roughly track phasic DA neuron firing. Unlike the striatum however, DAT concentrations in frontal cortex are low (e.g., Seamans and Robbins 2010). As a result, cortical DA levels change slowly—too slowly to track phasic DA activity. Even so, MODAL could be used in conjunction with almost any model of executive function that assigns a functional role to cortical DA levels. For example, Ashby et al. (2002) proposed a connectionist network model of creative problem solving that mapped loosely onto the anterior cingulate, prefrontal cortex, and head of the caudate nucleus. Although the model included little neuroanatomical detail, it made specific quantitative predictions about the effects of changing DA levels on cognitive flexibility and creative problem solving. No model of DA release was included, so MODAL could be used to fill this role.

Although MODAL could be used to predict changes in DA levels in any VTA DA target region and in virtually any task, it is important to note that how these changes affect behavior might be task and brain-region dependent. For example, in some tasks that depend on executive function, performance is an inverted U-shaped function of DA level. This includes creative problem solving and cognitive flexibility (Ashby et al. 1999; Cools and D'Esposito 2011; Cools and Robbins 2004; Cools 2006). Cools (2006) suggested that optimal levels of prefrontal DA facilitate the maintenance of stable representations, whereas optimal levels of striatal DA underlay cognitive flexibility. Therefore, changing global levels of DA can have different implications for task performance depending on local dynamics. Accordingly, although MODAL modulates DA input to these regions, the implications of changing global DA levels for behavior and performance will require region-specific models that consider local baseline DA levels, re-uptake mechanisms, receptor dynamics, and interactions between regions.

A more challenging goal is to extend MODAL to the DA neurons in the substantia nigra pars compacta (SNpc). The vSub → NAcc → VP pathway shown in Fig. 1 projects to VTA but not to SNpc, so accounting for changes in SNpc DA firing when environmental uncertainty or feedback contingency changes requires a different neuroanatomical model. This problem is complicated by recent evidence suggesting that despite many similarities, VTA and SNpc DA have dissociable roles (Keiflin et al. 2019). Furthermore, VTA and SNpc DA neurons project to different (but overlapping) targets. In particular, the dorsal striatum receives its DA projection almost exclusively from the SNpc (e.g., Smith and Kieval 2000). This is important because there is overwhelming evidence that procedural learning is mediated within the basal ganglia, and especially at cortical-striatal synapses in the dorsal striatum (e.g., Ashby and Ennis 2006; Houk et al. 1995; Mishkin et al. 1984; Willingham 1998). Therefore, to interface MODAL with models of the dorsal striatum and/or models of procedural learning, the model must be generalized to include the SNpc. One possibility is to model the spiraling architecture of the basal ganglia that enables activity in the ventral striatum (i.e., the NAcc) to influence the central striatum, which then influences the dorsolateral striatum (Takahashi et al. 2008; Haber et al. 2000; Belin and Everitt 2008). In fact, an existing actor-critic model of the basal ganglia already relies on this spiraling architecture (Takahashi et al. 2008).

Relation to RPE Models

The model proposed here describes how changes in some modulating variable affect the DA response to RPE. But note that the model makes no assumptions about the neural networks that compute RPE. Many models of these circuits have been proposed (Brown et al. 1999; Cohen et al. 2012; Contreras-Vidal and Schultz 1999; Eshel et al. 2015; Hazy and Frank 2010; Houk et al. 1995; Joel et al. 2002; Humphries and Prescott 2010; Kawato and Samejima 2007; Morita et al. 2012, 2013; O'Reilly et al. 2007; Salum et al. 1999; Schultz et al. 1997; Schultz 1998; Stuber et al. 2008; Sutton and Barto 1998; Tan and Bullock 2008; Vitay and Hamker 2014). MODAL does not generate behavior; therefore, rather than compute RPE using one of these proposed circuits, we chose to project hypothetical values of RPE (ranging from − 1 to + 1) to the VTA units via the PPTN or LH. This network architecture is consistent with accounts that midbrain DA neurons receive the signals necessary for computing RPE from upstream regions via the PPTN (Hong and Hikosaka 2014; Kobayashi and Okada 2007; Okada and Kobayashi 2013), LH (Tian and Uchida 2015; Hong et al. 2011; Matsumoto and Hikosaka 2007, 2009), and RMTN (Jhou et al. 2009).

We chose not to model the neural networks that compute RPE in an effort to provide a stronger test of the hypothesis that effects of the modulating variable on the DA response to RPE are mediated by a circuit that includes vSub, NAcc, and VP. Adding neural structures to compute RPE would increase the complexity of the model, thereby making it more difficult to attribute a success or failure of the overall network to one specific subnetwork. Even so, one advantage of the modeling approach followed here is the potential to develop "plug-and-play" models of different neural networks (Ashby 2018; Cantwell et al. 2017). Because MODAL is consistent with known neuroanatomy and neurophysiology, it should be possible to wire it into an existing similarly constrained model of the networks that compute RPE or networks that compute the modulating variable. This exercise is beyond the scope of the current application.

Relation to Existing Neural Models of Learning Rates

Several alternative neural accounts of how learning rates are modulated have been proposed. None of these include DA neurons however, and thus, to our knowledge, none can account for any of the neural data we considered in our benchmark tests. This section briefly discusses the more prominent of these alternative accounts, with a special emphasis on their relation to MODAL.

Bernacchia et al. (2011) reported single-unit recording results from monkeys that showed evidence that different neurons in ACC, PFC, and lateral intraparietal cortex are differentially sensitive to the time since the last reward. Based on these results, they proposed a neural network model in which a reservoir of such neurons could be used to dynamically alter learning rates, depending on how quickly environmental reward probabilities are changing.

Similarly, Farashahi et al. (2017) proposed that the ACC adjusts learning rates in response to environmental changes in reward probabilities via synaptic metaplasticity, which is a synaptic change that alters the plasticity of the synapse to future events, without altering the efficacy of current synaptic transmission. Specifically, they proposed that the ACC may be endowed with metaplastic synapses that can switch between strong and weak meta states, effectively changing the learning rate.

It is important to note, however, that neither of these proposals made any attempt to describe how the learning rate selected from the reservoir or computed in the ACC via metaplasticity, modulates neural plasticity in other brain networks. Thus, rather than competing with MODAL, these models could be viewed as candidate models for the network (or part of the network) that computes the modulating variable that serves as input to vSub in MODAL.

In contrast, a model that more directly competes with MODAL was proposed by Franklin and Frank (2015). According to this account, the pause duration of tonically active cholinergic neurons (TANs) in the striatum signals uncertainty and modulates learning by controlling the activity of the MSN population through a feedback loop. In this model, the TAN pauses are driven by input from the striatal MSNs. High entropy in the MSN population leads to long TAN pauses, which result in fast initial learning, whereas low entropy in the MSN population leads to short pauses, which result in slow initial learning. The result is a neural network that implements a dynamic learning rate that enables rapid learning after a reversal.

This is an interesting hypothesis that deserves further testing. Even so, it faces several significant challenges. First, Franklin and Frank (2015) acknowledged that they are unaware of any empirical support for the claim that TAN pause durations are modulated by MSN activity. Second, their model omits the strongest excitatory glutamatergic inputs to the TANs, which come from the caudal intralaminar nuclei of the thalamus (Cornwall and Phillipson 1988; Sadikot et al. 1992). Furthermore, simultaneous single-unit recordings from these thalamic neurons and from TANs show that thalamic activity is required for the TANs to pause (Matsumoto et al. 2001).

Third, there is evidence (acknowledged by Franklin and Frank 2015) that DA also modulates the duration of TAN pauses (Deng et al. 2007; Doig et al. 2014; Ding et al. 2010).

An alternative account of TAN activity was proposed by Ashby and Crossley (2011), who hypothesized that the main functional role of the TANs is to serve as a gate between cortex and the striatum. The TANs tonically inhibit cortical inputs to the striatum, so the default state of the gate is closed. However, environmental cues that signal reward cause the TANs to pause (via excitatory input from thalamus), which opens the gate and allows cortical-striatal plasticity. Furthermore, Crossley et al. (2013) proposed a model that included this role for the TANs, which protects cortical-striatal synapses when state-feedback contingency is low (e.g., as when the feedback is random), by eliminating the TAN pause to cues that formerly predicted reward (and thereby closing the gate). In this model, decreases in DA are necessary for the TANs to unlearn the pause response. Crossley et al. (2013) made no attempt to describe a neural circuit via which state-feedback contingency could modulate the amount of DA released, so MODAL could be combined with the Crossley et al. (2013) model to provide a more complete description of these contingency-related phenomena.

In summary, there are few true competitors to MODAL, but many models that could be combined with MODAL to produce a more powerful model than any that currently exists. Models in this latter class are of two types. One type, which includes the models of Bernacchia et al. (2011) and Farashahi et al. (2017), could be used to compute the value of the modulating variable that is the input to vSub in MODAL (see Fig. 1). Another type, which includes the Crossley et al. (2013) model, could use MODAL to compute the amount of DA released to feedback during each trial of some learning task. When combined in this way, MODAL would act as the critic, and the other model as the actor in an actor-critic architecture.

Neural Basis of Modulating Variables

MODAL proposes a neural account of how some modulating variable could affect the DA response to RPE and therefore learning rates in the brain. Many such variables have been proposed. MODAL does not require that the computation of all these putative variables are mediated by the same neural network, but it does require that the variable, whatever it is, is mediated by a network that sends a prominent projection to the vSub. Fortunately, almost all hypothesized modulating variables seem to meet this requirement (for a review of the numerous brain regions involved in coding uncertainty, see Soltani and Izquierdo 2019).

The vSub receives dense projections from the hippocampal CA1 subfield and from entorhinal cortex (Kerr et al. 2007), and these regions receive input from many areas of frontal cortex, including large portions of PFC, orbitofrontal cortex, and ACC (e.g., Gloor 1997). For example, entorhinal cortex receives almost all of its cortical inputs from polymodal association areas, including cingulate, orbitofrontal and parahippocampal cortices (Insausti et al. 1987; Jones and Witter 2007).

Almost all modulating variables are thought to depend on one or more of these regions. For example, the ACC seems to play a significant role in encoding volatility (Behrens et al. 2007), uncertainty (Rushworth and Behrens 2008), and valence-specific uncertainty (Monosov 2017). Activity in orbitofrontal cortex has been shown to correlate with uncertainty (Jo and Jung 2016; Neill and Schultz 2010) and additional evidence suggests that it may play a role in unexpected uncertainty and volatility (Riceberg and Shapiro 2012).

The encoding of unexpected uncertainty has been found in the posterior cingulate cortex, a portion of the postcentral gyrus and posterior insular cortex, the left middle temporal gyrus, the left hippocampus, and the locus coeruleus (Payzan-LeNestour et al. 2013). The encoding of estimation uncertainty has been found in the ACC extending to the posterior dorsomedial PFC, bilateral dorsolateral PFC, and a portion of the inferior parietal lobule (Payzan-LeNestour et al. 2013). The encoding of risk was found in the inferior frontal gyrus (Payzan-LeNestour et al. 2013; Huettel et al. 2005) and a portion of the lingual gyrus (Payzan-LeNestour et al. 2013), the adjacent anterior insula (Huettel et al. 2005; Preuschoff et al. 2008) and the ACC (Christopoulos et al. 2009). Preuschoff et al. (2008) found that activation in the insula encodes risk and risk prediction errors and Jo and Jung (2016) found that the anterior insula encodes signals related to reward uncertainty. Activity in the hippocampus has been found to correlate with uncertainty (Harrison et al. 2006; Vanni-Mercier et al. 2009; Strange et al. 2005). Furthermore, Payzan-LeNestour et al. (2013) noted the similarity between unexpected uncertainty and novelty detection and therefore the role played by the hippocampus in novelty detection may be relevant (Rutishauser et al. 2006). This proposal is particularly interesting when considering how the role of the hippocampus in mismatch detection may relate to the detection of changes in the environment (Kumaran and Maguire 2006). Dayan and Yu (2003) proposed that the effects of expected and unexpected uncertainty are mediated in cortex by acetylcholine and norepinephrine, respectively (Yu and Dayan 2005). This is relevant because Lipski and Grace (2013) showed that norepinephrine and locus coereleus activation can modulate the activity of neurons in vSub and Bortz and Grace (2018) showed that the modulation of VTA DA population size depends on cholinergic mechanisms in vSub. Additionally, lesions to the ventral striatum in monkeys have been shown to reduce learning rates in stochastic tasks, which is consistent with the role of NAcc in our model (Taswell et al. 2018). Finally, the medial septum has been shown to play a role in reversal learning in rats by controlling the size of active midbrain DA neurons and this effect was mediated via projections from medial septum to vSub (Bortz et al. 2019).

The architecture of MODAL implies that tonic DA encodes the learning rate. Therefore, our model is consistent with the proposal by Friston et al. (2012) suggesting that tonic DA encodes precision, that is, the learning rate in Bayesian models of learning under uncertainty (Mathys et al. 2011). Furthermore, using precision as a modulating variable in MODAL would enable our network to implement precision-weighted prediction errors.

Niv et al. (2007) proposed that tonic DA levels encode the average rate of reward in free-operant tasks. In the data used to test this model, pigeons and rats were trained in steady-state environments in which reward contingencies did not vary. Thus, the main modulating variables considered in this article, including uncertainty, volatility, and feedback contingency, are likely to have remained constant as well. As a result, more research is needed to distinguish between the average-reward-rate hypothesis and MODAL. Another possibility, however, is that average reward rate could be treated as a modulating variable that serves as input to MODAL.

The Benefit of Multiple Levels of Analysis and Future Research

We proposed an implementational-level model of how any of a variety of different modulating variables could control the gain on the DA response to RPE, and therefore implement dynamic learning rates. Although computational and algorithmic levels of analysis have been successful in accounting for behavioral phenomena, moving to the implementational level allows us to further constrain the models by the underlying neuroanatomy and neurophysiology, and brings to light questions that may not have been proposed at higher levels of analysis. Some questions that arise due to the implementational-level modeling are as follows: (1) What are the various computations encoded in cortical circuits that may act as inputs to MODAL? (2) In neural models of RPE, tonic DA levels often represent zero RPE; however, what happens when the tonic DA levels change? Note that MODAL predicts that increases in the value of the modulating variable should increase tonic concentrations of extracellular DA, even though it will not increase tonic firing rates in active DA neurons (however, note that local control mechanisms may also need to be considered; Berke 2018). To our knowledge, this prediction is untested and therefore should be investigated in detail. (3) What is the cellular or molecular mechanism that causes silent DA neurons to begin firing tonically? We modeled this transition by assuming there is variability in the resting state potential across the population of NAcc neurons. Topologically, this assumption caused some neurons to be further from a saddle-node bifurcation, and therefore to require more input current for the fixed points to undergo a saddle-node bifurcation. However, because the Izhikevich MSNs are phenomenological models of neural spiking, there are other possible mechanisms that could lead to similar dynamical behavior. Future research should test our proposed mechanism and investigate other possibilities.

Modeling at the implementational level has significant implications for disease states. Knowledge of the neurophysiology of disease can lead to hypotheses for models at the computational and algorithmic levels. For example, empirical evidence indicates that in schizophrenia, the DA system is in overdrive due to aberrant regulation of midbrain DA neurons by the vSub (Grace 2010). If the predictions of MODAL are considered, this kind of knowledge has implications for performance in a variety of behavioral tasks in people with schizophrenia.

Future research should extend our model upwards by investigating how MODAL could be integrated with the various circuits that have been proposed to monitor contextual and statistical aspects of the environment (e.g., as in Bernacchia et al. 2011; Farashahi et al. 2017). Greater specification of these circuits will enable us to take full advantage of the computational cognitive neuroscience approach by combining the circuits in a plug-and-play fashion. Finally, the model presented here was derived from neurobiological principles and meant to account for neurophysiological data and to serve as a foundation for the successful application of the model to behavioral data. Accordingly, future work should explore the application of the model to a variety of behavioral paradigms in which performance relies on DA levels, such as working memory, creative problem solving, reversal learning, task-set switching, category learning, and instrumental conditioning.

Funding Information This research was supported by NIH Grant 2R01MH063760.

Compliance with Ethical Standards

Competing Interests The authors declare no that they no competing interests.

References

Aminoff, E.M., Kveraga, K., Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, *17*(8), 379–390.

Ashby, F.G. (2018). Computational cognitive neuroscience. In Batchelder, W., Colonius, H., Dzhafarov, E., Myung, J. (Eds.) *New handbook of mathematical psychology, vol. 2 (pp. 223–270). New York*. New York: Cambridge University Press.

Ashby, F.G., & Crossley, M.J. (2011). A computational model of how cholinergic interneurons protect striatal-dependent learning. *Journal of Cognitive Neuroscience*, *23*(6), 1549–1566.

Ashby, F.G., & Ennis, J.M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1–36.

Ashby, F.G., & Helie, S. (2011). A tutorial on computational cognitive neuroscience: modeling the neurodynamics of cognition. *Journal of Mathematical Psychology*, *55*(4), 273–289.

Ashby, F.G., Isen, A.M., Turken, A. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review*, *106*(3), 529–550.

Ashby, F.G., Valentin, V.V., Turken, A.U. (2002). The effects of positive affect and arousal and working memory and executive attention: neurobiology and computational models. In Moore, S., & Oaksford, M. (Eds.) *Emotional cognition: from brain to behaviour* (pp. 245-287). Amsterdam: John Benjamins Publishing Company.

Ashby, F.G., & Vucovich, L.E. (2016). The role of feedback contingency in perceptual category learning. Journal of Experimental Psychology: Learning. *Memory, and Cognition*, *42*(11), 1731.

Bayer, H.M., & Glimcher, P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141.

Bayer, H.M., Lau, B., Glimcher, P.W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology*, *98*(3), 1428–1439.

Behrens, T.E., Woolrich, M.W., Walton, M.E., Rushworth, M.F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221.

Belin, D., & Everitt, B.J. (2008). Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. *Neuron*, *57*(3), 432–441.

Berke, J.D. (2018). What does dopamine mean? *Nature Neuroscience*, *21*(6), 787–793.

Bernacchia, A., Seo, H., Lee, D., Wang, X.-J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience*, *14*(3), 366–372.

Berridge, K.C. (2000). Reward learning: reinforcement, incentives, and expectations. In Medin, D. (Ed.) *Psychology of learning and motivation*, (Vol. 40 pp. 223–278): Elsevier.

Bland, A.R., & Schaefer, A. (2012). Different varieties of uncertainty in human decision-making. *Frontiers in Neuroscience*, *6*, 85.

Bortz, D.M., Gazo, K.L., Grace, A.A. (2019). The medial septum enhances reversal learning via opposing actions on ventral tegmental area and substantia nigra dopamine neurons. *Neuropsychopharmacology*, 1–9.

Bortz, D.M., & Grace, A.A. (2018). Medial septum differentially regulates dopamine neuron activity in the rat ventral tegmental area and substantia nigra via distinct pathways. *Neuropsychopharmacology*, *43*, 2093–2100.

Braganza, O., & Beck, H. (2018). The circuit motif as a conceptual tool for multilevel neuroscience. *Trends in Neurosciences*, *41*(3), 128–136.

Bromberg-Martin, E.S., Matsumoto, M., Hikosaka, O. (2010). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, *68*(5), 815–834.

Brown, J., Bullock, D., Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience*, *19*(23), 10502–10511.

Bush, R.R., & Mosteller, F. (1951). A model for stimulus generalization and discrimination. *Psychological Review*, *58*(6), 413–423.

Cantwell, G., Riesenhuber, M., Roeder, J.L., Ashby, F.G. (2017). Perceptual category learning and visual processing: an exercise in computational cognitive neuroscience. *Neural Networks*, *89*, 31–38.

Christopoulos, G.I., Tobler, P.N., Bossaerts, P., Dolan, R.J., Schultz, W. (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *Journal of Neuroscience*, *29*(40), 12574–12583.

Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*, *482*(7383), 85–88.

Contreras-Vidal, J.L., & Schultz, W. (1999). A predictive reinforcement model of dopamine neurons for learning approach behavior. *Journal of Computational Neuroscience*, *6*(3), 191–214.

Cools, R. (2006). Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neuroscience and Biobehavioral Reviews*, *30*(1), 1–23.

Cools, R., Clark, L., Owen, A.M., Robbins, T.W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, *22*(11), 4563–4567.

Cools, R., & D'Esposito, M. (2011). Inverted U-shaped dopamine actions on human working memory and cognitive control. *Biological Psychiatry*, *69*(12), e113–e125.

Cools, R., & Robbins, T.W. (2004). Chemistry of the adaptive mind. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical. Physical and Engineering Sciences*, *362*(1825), 2871–2888.

Cornwall, J., & Phillipson, O. (1988). Afferent projections to the parafascicular thalamic nucleus of the rat, as shown by the retrograde transport of wheat germ agglutinin. *Brain Research Bulletin*, *20*(2), 139–150.

Crossley, M.J., Ashby, F.G., Maddox, W.T. (2013). Erasing the engram: the unlearning of procedural skills. *Journal of Experimental Psychology: General*, *142*(3), 710–741.

Daw, N.D., & O'Doherty, J.P. (2014). Multiple systems for value learning. In P. W. Glimcher, & E. Fehr (Eds.) *Neuroeconomics: decision making and the brain, Second edition* (pp. 393–410). Amsterdam: Elsevier.

Dayan, P., & Abbott, L.F. (2001). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. Cambridge: MIT Press.

Dayan, P., Kakade, S., Montague, P.R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*(11s), 1218–1223.

Dayan, P., & Long, T. (1998). Statistical models of conditioning. In Jordan, M.I., Kearns, M.J., Solla, S.A. (Eds.) *Advances in neural information processing systems: Proceedings of the 1997 Conference* (pp. 117-123). Cambridge, MA: MIT Press.

Dayan, P., & Yu, A.J. (2003). Expected and unexpected uncertainty: ACh and NE in the neocortex. In Becker, S., Thrun, S., Obermayer, K. (Eds.) *Advances in neural information processing systems: Proceedings of the 2002 Conference* (pp. 173-180). Cambridge, MA: MIT Press.

Deng, P., Zhang, Y., Xu, Z.C. (2007). Involvement of I_h in dopamine modulation of tonic firing in striatal cholinergic interneurons. *Journal of Neuroscience*, *27*(12), 3148–3156.

Ding, J.B., Guzman, J.N., Peterson, J.D., Goldberg, J.A., Surmeier, D.J. (2010). Thalamic gating of corticostriatal signaling by cholinergic interneurons. *Neuron*, *67*(2), 294–307c.

Doig, N.M., Magill, P.J., Apicella, P., Bolam, J.P., Sharott, A. (2014). Cortical and thalamic excitation mediate the multiphasic responses of striatal cholinergic interneurons to motivationally salient stimuli. *Journal of Neuroscience*, *34*(8), 3101–3117.

Ermentrout, G.B. (1996). Type I membranes, phase resetting curves, and synchrony. *Neural Computation*, *8*(5), 979–1001.

Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., Uchida, N. (2015). Arithmetic and local circuitry underlying dopamine prediction errors. *Nature*, *525*(7568), 243–246.

Fabbricatore, A.T., Ghitza, U.E., Prokopenko, V.F., West, M.O. (2009). Electrophysiological evidence of mediolateral functional dichotomy in the rat accumbens during cocaine self-administration: tonic firing patterns. *European Journal of Neuroscience*, *30*(12), 2387–2400.

Faget, L., Osakada, F., Duan, J., Ressler, R., Johnson, A.B., Proudfoot, J.A., Hnasko, T.S. (2016). Afferent inputs to neurotransmitter-defined cell types in the ventral tegmental area. *Cell reports*, *15*(12), 2796–2808.

Fanselow, M.S., & Dong, H.W. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron*, *65*(1), 7–19.

Farashahi, S., Donahue, C.H., Khorsand, P., Seo, H., Lee, D., Soltani, A. (2017). Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron*, *94*(2), 401–414.

Franklin, N.T., & Frank, M.J. (2015). A cholinergic feedback circuit to regulate striatal population uncertainty and optimize reinforcement learning. *Elife*, *4*.

Friston, K.J., Shiner, T., FitzGerald, T., Galea, J.M., Adams, R., Brown, H., Bestmann, S. (2012). Dopamine, affordance and active inference. *PLoS Computational Biology*, 8(1).

Gloor, P. (1997). *The temporal lobe and limbic system*. New York: Oxford University Press.

Grace, A.A. (2010). Dopamine system dysregulation by the ventral subiculum as the common pathophysiological basis for schizophrenia psychosis, psychostimulant abuse, and stress. *Neurotoxicity Research*, *18*(3-4), 367–376.

Grace, A.A., & Bunney, B.S. (1983). Intracellular and extracellular electrophysiology of nigral dopaminergic neurons-1. Identification and characterization. *Neuroscience*, *10*(2), 301–315.

Grace, A.A., Floresco, S.B., Goto, Y., Lodge, D.J. (2007). Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends in Neurosciences*, *30*(5), 220–227.

Haber, S.N. (2016). Corticostriatal circuitry. *Dialogues in Clinical Neuroscience*, *18*(1), 7.

Haber, S.N., Fudge, J.L., McFarland, N.R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience*, *20*(6), 2369–2382.

Harrison, L.M., Duggins, A., Friston, K.J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks*, *19*(5), 535–546.

Hart, A.S., Rutledge, R.B., Glimcher, P.W., Phillips, P.E. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *Journal of Neuroscience*, *34*(3), 698–704.

Hazy, T.E., & Frank, M.J. (2010). O'Reilly, R. C Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience and Biobehavioral Reviews*, *34*(5), 701–720.

Hong, S., & Hikosaka, O. (2014). Pedunculopontine tegmental nucleus neurons provide reward, sensorimotor, and alerting signals to midbrain dopamine neurons. *Neuroscience*, *282*, 139–155.

Hong, S., Jhou, T.C., Smith, M., Saleem, K.S., Hikosaka, O. (2011). Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. *Journal of Neuroscience*, *31*(32), 11457–11471.

Horvitz, J.C. (2002). Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. *Behavioural Brain Research*, *137*(1-2), 65–74.

Houk, J., Adams, J., Barto, A. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In Davis, J.L., Beiser, D.G., Houk J.C. (Eds.) *Models of information processing in the basal ganglia* (pp. 249-270). Cambridge: MIT Press.

Huettel, S.A., Song, A.W., McCarthy, G. (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *Journal of Neuroscience*, *25*(13), 3304–3311.

Humphries, M.D., & Prescott, T.J. (2010). The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Progress in Neurobiology*, *90*(4), 385–417.

Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *Elife*, *5*, e18073.

Insausti, R., Amaral, D., Cowan, W. (1987). The entorhinal cortex of the monkey: II. Cortical afferents. *Journal of Comparative Neurology*, *264*(3), 356–395.

Izhikevich, E.M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, *14*(6), 1569–1572.

Izhikevich, E.M. (2007). *Dynamical systems in neuroscience*. Cambridge, CA: MIT Press.

Jacobs, J., Kahana, M.J., Ekstrom, A.D., Mollison, M.V., Fried, I. (2010). A sense of direction in human entorhinal cortex. *Proceedings of the National Academy of Sciences*, *107*(14), 6487–6492.

Jhou, T.C., Fields, H.L., Baxter, M.G., Saper, C.B., Holland, P.C. (2009). The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron*, *61*(5), 786–800.

Jo, S., & Jung, M.W. (2016). Differential coding of uncertain reward in rat insular and orbitofrontal cortex. *Scientific Reports*, *6*, 24085.

Joel, D., Niv, Y., Ruppin, E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, *15*(4-6), 535–547.

Jones, B.F., & Witter, M.P. (2007). Cingulate cortex projections to the parahippocampal region and hippocampal formation in the rat. *Hippocampus*, *17*(10), 957–976.

Kawato, M., & Samejima, K. (2007). Efficient reinforcement learning: computational theories, neuroscience and robotics. *Current Opinion in Neurobiology*, *17*(2), 205–212.

Keiflin, R., Pribut, H.J., Shah, N.B., Janak, P.H. (2019). Ventral tegmental dopamine neurons participate in reward identity predictions. *Current Biology*, *29*(1), 93–103.

Kerr, K.M., Agster, K.L., Furtak, S.C., Burwell, R.D. (2007). Functional neuroanatomy of the parahippocampal region: the lateral and medial entorhinal areas. *Hippocampus*, *17*(9), 697–708.

Kobayashi, Y., & Okada, K. (2007). Reward prediction error computation in the pedunculopontine tegmental nucleus neurons. *Annals of the New York Academy of Sciences*, *1104*(1), 310–323.

Kumaran, D., & Maguire, E.A. (2006). An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS Biology*, *4*(12), e424.

Lipski, W.J., & Grace, A.A. (2013). Activation and inhibition of neurons in the hippocampal ventral subiculum by norepinephrine and locus coeruleus stimulation. *Neuropsychopharmacology*, *38*(2), 285.

Liu, X., Hairston, J., Schrier, M., Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–1236.

Lodge, D.J., & Grace, A.A. (2006). The hippocampus modulates dopamine neuron responsivity by regulating the intensity of phasic neuron activation. *Neuropsychopharmacology*, *31*(7), 1356–1361.

Maia, T.V. (2009). Reinforcement learning, conditioning, and the brain: successes and challenges. *Cognitive, Affective, and Behavioral Neuroscience*, *9*(4), 343–364.

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. New York: Freeman.

Mathys, C., Daunizeau, J., Friston, K.J., Stephan, K.E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39.

Matsumoto, M., & Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, *447*(7148), 1111–1115.

Matsumoto, M., & Hikosaka, O. (2009). Representation of negative motivational value in the primate lateral habenula. *Nature Neuroscience*, *12*(1), 77–84.

Matsumoto, N., Minamimoto, T., Graybiel, A.M., Kimura, M. (2001). Neurons in the thalamic CM-Pf complex supply striatal neurons with information about behaviorally significant sensory events. *Journal of Neurophysiology*, *85*(2), 960–976.

Mishkin, M., Malamut, B., Bachevalier, J. (1984). Memories and habits: two neural systems. In Lynch, G., McGaugh, J.L., Weinberger, N.M. (Eds.) *Neurobiology of human learning and memory* (pp. 65-77). New York: Guilford Press.

Monosov, I.E. (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. *Nature Communications*, *8*(1), 134.

Montague, P.R., Dayan, P., Sejnowski, T.J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947.

Morita, K., Morishima, M., Sakai, K., Kawaguchi, Y. (2012). Reinforcement learning: computing the temporal difference of values via distinct corticostriatal pathways. *Trends in Neurosciences*, *35*(8), 457–467.

Morita, K., Morishima, M., Sakai, K., Kawaguchi, Y. (2013). Dopaminergic control of motivation and reinforcement learning: a closed-circuit account for reward-oriented behavior. *Journal of Neuroscience*, *33*(20), 8866–8890.

Niv, Y., Daw, N.D., Joel, D., Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520.

Okada, K., & Kobayashi, Y. (2013). Reward prediction-related increases and decreases in tonic neuronal activity of the pedunculopontine tegmental nucleus. *Frontiers in Integrative Neuroscience*, *7*, 36.

Neill, M., & Schultz, W. (2010). Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron*, *68*(4), 789–800.

O'Reilly, R.C., Frank, M.J., Hazy, T.E., Watz, B. (2007). PVLV: the primary value and learned value Pavlovian learning algorithm. *Behavioral Neuroscience*, *121*(1), 31–49.

Payzan-LeNestour, E., & Bossaerts, P. (2011). Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Computational Biology*, *7*(1), e1001048.

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., O'Doherty, J.P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron*, *79*(1), 191–201.

Pickering, A.D., & Pesola, F. (2014). Modeling dopaminergic and other processes involved in learning from reward prediction error: contributions from an individual differences perspective. *Frontiers in Human Neuroscience*, *8*, 740.

Preuschoff, K., & Bossaerts, P. (2007). Adding prediction risk to the theory of reward learning. *Annals of the New York Academy of Sciences*, *1104*(1), 135–146.

Preuschoff, K., Quartz, S.R., Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience*, *28*(11), 2745–2752.

Quintero, E., Diaz, E., Vargas, J.P., de la Casa, G., Lopez, J.C. (2011). Ventral subiculum involvement in latent inhibition context specificity. *Physiology and Behavior*, *102*(3-4), 414–420.

Rall, W. (1967). Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *Journal of Neurophysiology*, *30*(5), 1138–1168.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black, A.H., & Prokasy, W.F. (Eds.) *Classical conditioning II: current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Riceberg, J.S., & Shapiro, M.L. (2012). Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. *Journal of Neuroscience*, *32*(46), 16402–16409.

Root, D.H., Fabbricatore, A.T., Pawlak, A.P., Barker, D.J., Ma, S., West, M.O. (2012). Slow phasic and tonic activity of ventral pallidal neurons during cocaine self-administration. *Synapse*, *66*(2), 106–127.

Rushworth, M.F., & Behrens, T.E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*(4), 389–397.

Rutishauser, U., Mamelak, A.N., Schuman, E.M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron*, *49*(6), 805–813.

Sadikot, A., Parent, A., Francois, C. (1992). Efferent connections of the centromedian and parafascicular thalamic nuclei in the squirrel monkey: a PHA-L study of subcortical projections. *Journal of Comparative Neurology*, *315*(2), 137–159.

Salum, C., da Silva, A.R., Pickering, A. (1999). Striatal dopamine in attentional learning: a computational model. *Neurocomputing*, *26*, 845–854.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27.

Schultz, W., Dayan, P., Montague, P.R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Seamans, J.K., & Robbins, T.W. (2010). Dopamine modulation of the prefrontal cortex and cognitive function. In Neve, K.A. (Ed.) *The dopamine receptors*. 2nd edn. (pp. 373-398). New York: Springer.-.

Sesack, S.R., & Grace, A.A. (2010). Cortico-basal ganglia reward network: microcircuitry. *Neuropsychopharmacology*, *35*(1), 27–47.

Smith, Y., & Kieval, J.Z. (2000). Anatomy of the dopamine system in the basal ganglia. *Trends in Neurosciences*, *23*, S28–S33.

Soltani, A., & Izquierdo, A. (2019). Adaptive learning under expected and unexpected uncertainty. *Nature Reviews Neuroscience*, *20*(10), 635–644.

Strange, B.A., Duggins, A., Penny, W., Dolan, R.J., Friston, K.J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Networks*, *18*(3), 225–230.

Stuber, G.D., Klanker, M., De Ridder, B., Bowers, M.S., Joosten, R.N., Feenstra, M.G., Bonci, A. (2008). Reward predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science*, *321*(5896), 1690–1692.

Sutton, R.S. (1992). Adapting bias by gradient descent: an incremental version of delta-bar-delta. In *Proceedings of the tenth national conference on artificial intelligence* (pp. 171–176). Cambridge: MIT Press.

Sutton, R.S., & Barto, A.G. (1998). *Reinforcement learning: an introduction Cambridge*. MA: MIT Press.

Takahashi, Y.K., Langdon, A.J., Niv, Y., Schoenbaum, G. (2016). Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. *Neuron*, *91*(1), 182–193.

Takahashi, Y.K., Schoenbaum, G., Niv, Y. (2008). Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. *Frontiers in neuroscience*, *2*, 14.

Tan, C.O., & Bullock, D. (2008). A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. *Journal of Neuroscience*, *28*(40), 10062–10074.

Taswell, C.A., Costa, V.D., Murray, E.A., Averbeck, B.B. (2018). Ventral striatum's role in learning from gains and losses. *Proceedings of the National Academy of Sciences*, *115*(52), E12398–E12406.

Tian, J., & Uchida, N. (2015). Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. *Neuron*, *87*(6), 1304–1316.

Vanni-Mercier, G., Mauguiere, F., Isnard, J., Dreher, J.-C. (2009). The hippocampus codes the uncertainty of cue-outcome associations: an intracranial electrophysiological study in humans. *Journal of Neuroscience*, *29*(16), 5287–5294.

Van Rossum, G., & Drake, F.L. (2011). The Python language reference manual. Network Theory Ltd.

Vitay, J., & Hamker, F.H. (2014). Timing and expectation of reward: a neuro-computational model of the afferents to the ventral tegmental area. *Frontiers in Neurorobotics*, *8*, 4.

Watabe-Uchida, M., Zhu, L., Ogawa, S.K., Vamanrao, A., Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron*, *74*(5), 858–873.

Willingham, D.B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review*, *105*, 558–584.

Yu, A.J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.