RESEARCH ARTICLE | *Higher Neural Functions and Behavior*

Adapting the flow of time with dopamine

^(D) John G. Mikhael¹ and Samuel J. Gershman²

¹Program in Neuroscience and MD-PhD Program, Harvard Medical School, Boston, Massachusetts; and ²Center for Brain Science and Department of Psychology, Harvard University, Cambridge, Massachusetts

Submitted 4 December 2018; accepted in final form 20 February 2019

Mikhael JG, Gershman SJ. Adapting the flow of time with dopamine. J Neurophysiol 121: 1748-1760, 2019. First published March 13, 2019; doi:10.1152/jn.00817.2018.-The modulation of interval timing by dopamine (DA) has been well established over decades of research. The nature of this modulation, however, has remained controversial: Although the pharmacological evidence has largely suggested that time intervals are overestimated with higher DA levels, more recent optogenetic work has shown the opposite effect. In addition, a large body of work has asserted DA's role as a "reward prediction error" (RPE), or a teaching signal that allows the basal ganglia to learn to predict future rewards in reinforcement learning tasks. Whether these two seemingly disparate accounts of DA may be related has remained an open question. By taking a reinforcement learning-based approach to interval timing, we show here that the RPE interpretation of DA naturally extends to its role as a modulator of timekeeping and furthermore that this view reconciles the seemingly conflicting observations. We derive a biologically plausible, DA-dependent plasticity rule that can modulate the rate of timekeeping in either direction and whose effect depends on the timing of the DA signal itself. This bidirectional update rule can account for the results from pharmacology and optogenetics as well as the behavioral effects of reward rate on interval timing and the temporal selectivity of striatal neurons. Hence, by adopting a single RPE interpretation of DA, our results take a step toward unifying computational theories of reinforcement learning and interval timing.

NEW & NOTEWORTHY How does dopamine (DA) influence interval timing? A large body of pharmacological evidence has suggested that DA accelerates timekeeping mechanisms. However, recent optogenetic work has shown exactly the opposite effect. In this article, we relate DA's role in timekeeping to its most established role, as a critical component of reinforcement learning. This allows us to derive a neurobiologically plausible framework that reconciles a large body of DA's temporal effects, including pharmacological, behavioral, electrophysiological, and optogenetic.

dopamine; interval timing; reinforcement learning; reward prediction error

INTRODUCTION

The ability to accurately estimate time is crucial for survival and is critically intertwined with reinforcement learning (RL). To adequately learn the properties of its environment and respond to them as they change, an animal must learn the temporal information associated with both its environment and its own behavior (Allman et al. 2014; Buhusi and Meck 2005; Matthews and Meck 2014).

Over the past several decades, dopamine (DA) has emerged as a neural substrate with a central role in both RL and interval timing. In RL tasks, midbrain DA neurons communicate a "reward prediction error" (RPE), or the signed difference between received and expected reward (Schultz et al. 1997). This RPE serves as a teaching signal that instructs basal ganglia (BG) circuitry to update its estimates of incoming rewards until those rewards are well predicted (Eshel et al. 2015; Glimcher 2011; Niv and Schoenbaum 2008; Schultz et al. 1997; Steinberg et al. 2013). Hence, delivery of an unexpected reward elicits a DA burst, omission of an expected reward elicits a DA dip, and a fully predicted reward elicits no DA response. Subsequent work has shown that DA maps to the RPE term of temporal difference learning models particularly well (Schultz 2007), and we review this class of algorithms below.

On the other hand, decades of research have implicated DA and the BG in interval timing (Buhusi and Meck 2005). Timing dysfunction has been well documented in diseases of the DA system and the BG such as Parkinson's disease (Malapani et al. 1998), Huntington's disease (Rowe et al. 2010), and schizophrenia (Elvevåg et al. 2003). In Parkinson's disease, dysfunction in timing is often ameliorated with DA agonists (Artieda et al. 1992; Jahanshahi et al. 2010; Malapani et al. 1998). This dysfunction has been recapitulated in healthy subjects pharmacologically (Arushanyan et al. 2003; Maricq and Church 1983) and through transgenic (Ward et al. 2009) and optogenetic (Soares et al. 2016) manipulation of the DA system in mouse models as well as by direct lesion of striatum in rats (Meck 2006).

Unlike its role as an RPE in RL tasks, however, a conceptual model of DA in timekeeping has been elusive, in part because of the recent emergence of conflicting evidence regarding its role. The pharmacological evidence, with notable exceptions, has tended to show that DA agonists result in behaviors consistent with a faster "internal clock" and DA antagonists with a slower clock (Cheng et al. 2007; Lake and Meck 2013; Maricq et al. 1981; Maricq and Church 1983; Meck 1986). On the other hand, recent photometric recordings and optogenetic manipulations have shown exactly the opposite effect (Soares et al. 2016). How to explain these seemingly conflicting results and, in fact, why DA would serve a role in modulating timekeeping in the first place remain open questions. Furthermore, whether and how the two seemingly disparate roles of

Address for reprint requests and other correspondence: J. G. Mikhael, Northwest Laboratory 295.06, 52 Oxford St., Cambridge, MA 02138 (e-mail: john_mikhael@hms.harvard.edu).

Downloaded from www.physiology.org/journal/jn at Harvard Univ (206.253.207.235) on May 1, 2019.

DA in RL and timekeeping may be related remains unknown (Gershman et al. 2014; Petter et al. 2018).

We show in this article that viewing interval timing through an RL lens allows us to reconcile the conflicting evidence on the role of DA in timing tasks. To do so, we derive from first principles a local plasticity rule that describes how DA allows cortico-striatal circuits to learn the temporal structure of reward. This rule accounts for how the DA signal modulates the internal clock and shows that it can produce either a speeding or a slowing effect, depending on when during the task it occurs.

METHODS

For all simulations of experimental results, we have chosen $\mu_d = d$ for $d \in \{1, 2, ..., 80\}$, $\gamma = 0.9$, $\sigma = 10$, $\mathcal{T} = 40$, $\tau = t^{0.7}$, and $\alpha_{\eta} = 0.1$, unless otherwise stated below.

Simulation of behavioral results. Reinforcement density (RD) is the inverse of trial length. We have initialized baseline reward time to be T = 30. For high and low bias, we have selected T = 25 and T = 35, respectively, with subsequent return to the baseline condition T = 30 (see Fig. 5).

Simulation of electrophysiological results. By visual inspection, we set $\sigma = 30$ and $\tau = t^{0.6}$ (see Fig. 6).

Simulation of pharmacological results. Update over the course of a single trial occurred according to the approximation in Eq. 7. To visually match the deviation of responses in medicated conditions from those of placebo, we have arbitrarily scaled the total update by $\alpha_{\eta} = 20$ (note that α_{η} need not be <1) and set $\sigma = 5$ and $T_{\text{max}} = 35$ for the placebo condition. For the temporal estimates, we have plotted $x_{t=7}$ and $x_{t=17}$. We take the end point of temporal reproduction to occur at the peak of \hat{V}_{τ} . For administration of DA agonists, we have chosen a gain modulation of 5 for positive RPEs and 0.2 for negative RPEs; for administration of DA antagonists, we have chosen a gain modulation of 0.2 for positive RPEs and 5 for negative RPEs. Finally, we have fixed the veridical time estimate to correspond to the baseline condition of unitary modulation of all RPEs (see Fig. 7).

Simulation of optogenetic results. To simulate the results of Soares et al. (2016), we have chosen $\alpha_{\eta} = 10$ (α_{η} need not be <1). A mean of 1.5 and a temperature parameter of 0.2 in the softmax learning rule (Luce 1959; Shepard 1958) were used to generate the choices. For activation and inhibition, we have arbitrarily chosen $\delta_{\tau} = +1$ and $\delta_{\tau} = -1$, respectively (see Fig. 8). For the results of Toda et al. (2017), we have chosen $\alpha_{\eta} = 7$ and $T_{\text{max}} = 20$ and set $\delta_{\tau} = +1$ (see Fig. 9).

Source code. Source code for all simulations can be found at https://www.github.com/jgmikhael/flowoftime.

RESULTS

As our model endows interval timing frameworks with an RL architecture, we begin this section with a brief review of temporal difference learning algorithms.

Temporal difference learning in cortico-basal ganglia circuitry. Perhaps the most successful account of DA function in the BG posits that phasic DA activity reports the RPE, or the difference between the received and expected reward (Glimcher 2011; Schultz 2007; Schultz et al. 1997). This account has been formalized in terms of the temporal difference (TD) learning algorithm (Sutton 1988), which we briefly review here.

For ease of exposition, we omit actions and assume that the agent traverses a series of "states" according to a Markov process. A state corresponds to some combination of relevant contextual cues, or features, that aid in predicting future rewards. For example, a state could denote proximity to a known food source or time until reward delivery. In the TD framework, the value of being in a state is defined as the expected discounted sum of current and future rewards:

$$V_t = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k}\right] \tag{1}$$

where t denotes time and indexes states, r_t is the reward received at time t, and $\gamma \in [0,1)$ is a discount factor that decreases the weights of later rewards. Following previous work (Ludvig et al. 2008, 2012; Schultz et al. 1997), the agent estimates V_t by learning a linear weighting of the features:

$$\hat{V}_t = \sum_d w_d x_{d,t} \tag{2}$$

where $x_{d,t}$ is the *d*th feature at time *t*, and this feature is weighted by w_d . The weights are updated by gradient ascent to reduce the discrepancy between V_t and its estimate \hat{V}_t :

$$v_d^{(t+1)} = w_d^{(t)} + \alpha \delta_t^{(t)} \nabla_{w_d} \hat{V}_t^{(t)}$$
(3)

where $\alpha \in [0,1)$ denotes the learning rate, the superscript represents the learning step, $\nabla_{w_d} \hat{V}_t^{(t)} = x_{d,t}$ is the gradient of $\hat{V}_t^{(t)}$ with respect to the weight w_d and δ_t is the RPE:

$$\delta_t = r_t + \gamma \hat{V}_{t+1} - \hat{V}_t \tag{4}$$

For an intuition, the received reward r_t , together with the discounted expected value at the next step $\gamma \hat{V}_{t+1}$, can be thought of as a single sample of value V_t . RPE is then the difference between this sample and the current estimate \hat{V}_t .

In mapping the TD framework onto cortico-BG circuitry, the features $x_{d,t}$ represent cortical inputs to striatum, the weights w_d are encoded by the strengths of the cortico-striatal synapses (Houk et al. 1995; Montague et al. 1996), the error signal δ_t is communicated by the firing of midbrain DA neurons (Eshel et al. 2015; Glimcher 2011; Niv and Schoenbaum 2008; Schultz et al. 1997; Steinberg et al. 2013), and the estimated value \hat{V}_t constitutes the output of the BG (Ratcliff and Frank 2012).

For a timekeeping task, a natural feature set arises from the recent experimental observation of "time cells," or neurons that are sequentially activated during the timed interval and that predict the animal's estimate of the interval (Mello et al. 2015; Wang et al. 2018). Remarkably, when the length of the required interval duration was experimentally manipulated, the response profiles of these cells rescaled to fit the newly timed duration as the animal learned the updated parameters of the task. This rescaling is an important attribute of our model, which we describe next.

Model description. We adopt the approximation architecture outlined in Eq. 2, where the features $x_{d,\tau}$ and estimated value \hat{V}_{τ} are indexed by subjective time τ . The features are encoded by time cells, and each time cell d is preferentially tuned to a subjective time μ_d , such that higher values of d correspond to cells that respond later. For simplicity, we take the features $x_{d,\tau}$ to be Gaussian in shape, centered at μ_d , and with tuning width σ , i.e.,

$$x_{d,\tau} = \exp\left[-\frac{1}{2}\left(\frac{\tau - \mu_d}{\sigma}\right)^2\right]$$
(5)

One simple neural circuit that implements this architecture is diagrammed in the bottom three layers of Fig. 1, where time



Fig. 1. Model architecture. The top layer denotes a compressive representation t' of objective time t. t' maps onto subjective time τ , weighted by the scaling factor η . Each time cell d is preferentially tuned to a time μ_d and responds with activation $x_{d,\tau}$. The sum of the features $x_{d,\tau}$, weighted by w_d , gives estimated value \hat{V}_{τ} .

cells are fed subjective time τ and their activations are weighted and summed to give the estimated value \hat{V}_{τ} , as described in *Temporal difference learning in cortico-basal ganglia circuitry*.

Let us now introduce our key addition to the linear approximation architecture, which derives from the observation that time cell responses rescale to reflect the properties of the external environment (Mello et al. 2015; Wang et al. 2018), as mentioned above. We implement this experimental finding with the parsimonious addition of a scaling factor η between t' and τ , where t' represents a compressive function of objective time t (see appendix for a comparison with alternative implementations) and t' and τ are represented upstream of the time cells, presumably in cortex (Wang et al. 2018). Hence, η controls the mapping of objective time onto subjective time, where a higher η means a faster clock. In keeping with conventional terminology (Gibbon et al. 1997; Treisman 1963; Zakay and Block 1997), we also refer to the clock as a "pacemaker," and hence refer to the parameter η as the pacemaker rate. The novelty of our account is the idea that the pacemaker rate, known to be modulated by DA as discussed in INTRODUCTION, can be viewed as a parameter of the function approximation architecture. This parameter is thus updated by gradient ascent:

$$\eta^{(t+1)} = \eta^{(t)} + \alpha_{\eta} \delta_{\tau}^{(t)} \frac{\tau}{\eta^{(t)}} \dot{V}_{\tau}^{(t)}$$
(6)

where α_{η} is the learning rate, $\nabla_{\eta}\hat{V} = \frac{\partial\hat{V}}{\partial\tau}\frac{\partial\tau}{\partial\eta} = \frac{\tau}{\eta}\hat{V}_{\tau}$ is the

gradient of estimated value with respect to η , and \hat{V}_{τ} denotes the time derivative of \hat{V}_{τ} .

We have assumed a constant tuning width σ for all features against subjective time, but note that because subjective time is a compressive function of objective time, the widths of the features increase with the passage of objective time, as shown in Fig. 2. This compression will not be necessary for our results below but better reflects the universal observation that subjective estimates of elapsed duration become (linearly) more corrupted as the duration increases, a phenomenon known as the scalar property (Gibbon 1977). It also facilitates comparison of our features with those of the microstimulus model, which are also overlapping Gaussians of increasing tuning width (Ludvig et al. 2008), and the learning-to-time model (LeT) from the interval timing literature, which posits a sequential activation of behavioral states with associated learnable weights (Machado 1997).

A bidirectional update rule governs η . To examine the update rule in Eq. 6 more closely, let us restrict our analysis to the case of a single reward of magnitude 1 delivered after a delay of objective time T, which corresponds to subjective time \mathcal{T} . In this case, Eq. 1 reduces to $V_{\tau} = \gamma^{\mathcal{T}-\tau}$ over the delay period (Fig. 3, top, dashed gray curve). Importantly, with an overlapping feature set, our architecture can never learn V_{τ} exactly. Instead, due to "smearing" across states, the estimate \hat{V}_{τ} will resemble a smeared version of V_{τ} (Fig. 3, top, solid black curve). This in turn will result in an RPE that is not fully reducible even with extensive learning, as has been theoretically predicted (Ludvig et al. 2008) and experimentally suggested (Kobayashi and Schultz 2008). For instance, internal timing noise, which may be implemented by an overlapping feature set, precludes the exact time of reward from being perfectly estimated; hence, a small amount of "surprise" (or nonzero RPE) will be experienced even if the reward is delivered on time (Fig. 3, bottom, solid black curve). In addition, owing to the asymmetric shape of V_t about time T, the peak of the estimate \hat{V}_{τ} occurs slightly before the true time of reward delivery T. We return to this point in a subsequent result.

There are two critical observations to be made about the update rule in Eq. 6. First, the synaptic weight encoding η has



Fig. 2. Effect of scaling η on activation of time cells. When measured against objective time *t*, higher η leads to more compressed time cell activations. Here, the same η for all features is learned. By *Eq.* 5, it is straightforward to show that all features move in tandem with a constant coefficient of variation, and no crossovers are produced during rescaling (appendix).



Fig. 3. Effect of state uncertainty on value estimation and reward prediction error (RPE). With a perfectly learned value function (*top*, gray), δ_{τ} reduces to zero throughout the entire trial duration (*bottom*, gray). With state uncertainty, implemented by an overlapping feature set, value cannot be estimated perfectly (*top*, black), and δ_{τ} is nonzero even after extensive learning (*bottom*, black). (Figure for illustration only. For smaller *T* or larger γ , the initial phasic RPE can be larger than RPE at reward time, illustrated here at t = 20 and t = 75, respectively.)

direct access to τ from its postsynaptic neuron (or its presyn-

aptic neuron, as $\tau = \eta t'$), to δ_{τ} from the DA signal, and to V_{τ} from projections from the BG output nuclei (substantia nigra pars reticulata and globus pallidus interna) via the BG-thalamo-cortical loop (Utter and Basso 2008). Therefore, the learning rule for η is local. The update rule itself is a form of differential Hebbian learning (Klopf 1988; Kosko 1986), whereby changes in presynaptic activity are correlated with postsynaptic activity to determine the extent of update. Roberts (1999) has shown that this type of learning rule can be implemented by spike-timing-dependent plasticity (see also Rao and Sejnowski 2001). This update is effectively gated by the RPE, giving rise to a form of three-factor learning rule similar to what has been proposed to underlie plasticity at cortico-striatal synapses (Reynolds and Wickens 2002).

Second, it follows from Eq. 6 that the update of η produces a bidirectional learning rule, as illustrated in Fig. 4. Conceptually, just as the weights w_d will increase or decrease to allow \hat{V}_{τ} to best approximate V_{τ} , so too will η . Whereas the weights w_d will change the magnitude of estimated value at different points of time, changes in η result in rescaling of estimated value against objective time. Hence, for η a reward that is delivered earlier than expected indicates that estimated value is not compressed enough, so η must increase. On the other hand, a reward that is delivered later than expected indicates that estimated value is too compressed, so η must decrease. For an intuition of how our derived update rule implements this

bidirectionality, note that this rule relies on the product of \hat{V}_{τ} and δ_{τ} . After learning, \hat{V}_{τ} takes a monotonically increasing (and roughly convex) shape leading up to reward time \mathcal{T} , closely approximating $V_{\tau} = \gamma^{\mathcal{T}-\tau}$, and then decreases quickly, but not instantaneously, to zero (Fig. 4, black curve, and Fig. 3, *top*, black curve). This more gradual decrease is due to the overlapping nature of our features, as late features accrue positive weights due to their contribution to the estimated value leading up to reward. The shape of \hat{V}_{τ} implies that its derivative

 \hat{V}_{τ} will be >0 before reward delivery, ~0 zero at reward delivery, and <0 during a window afterward. Therefore, early reward will result in $\Delta \eta \sim \dot{V}_{\tau} \delta_{\tau} > 0$ and thus an increase in η and compression of the features against objective time. On the other hand, late reward will result in a negative RPE over $\tau < \mathcal{T}$ when $\dot{V}_{\tau} > 0$, followed by a positive RPE when reward is delivered, at which point $\dot{V}_{\tau} < 0$; hence, $\Delta \eta \sim \dot{V}_{\tau} \delta_{\tau} < 0$ over both domains, so that η decreases and the features expand.

In summary, Eq. 6 represents a local update rule of the pacemaker rate η , where DA can facilitate either an increase or a decrease in η to allow the estimated value \hat{V}_{τ} to better approximate the true value V_{τ} .

Finally, having examined the role of phasic DA in communicating RPEs to update η , let us now consider how tonic DA levels influence the learning of η . Numerous studies have shown that higher levels of tonic DA enhance learning from positive RPEs, whereas low tonic DA levels enhance learning from negative RPEs (Cools et al. 2009; Frank et al. 2007, 2004; Shiner et al. 2012; Smittenaar et al. 2012). Assuming temporal reproduction occurs by timing until the peak of the estimated

value function, over which $\hat{V}_{\tau} > 0$, it follows from Eq. 6 that higher tonic DA levels will result in a bias toward increases of η . Hence, the pacemaker will accelerate and the features will be compressed against objective time. Similarly, low tonic DA levels will decelerate the pacemaker and expand the features.

Relationship with experimental data. A large body of experimental work in the interval timing literature is concerned with the modulation of timekeeping in healthy states and in disease. For instance, even in healthy states, timekeeping is biased by factors such as attention (Coull et al. 2004), motivation (Gable



Fig. 4. Bidirectional learning rule for η . Because of feature overlap, the learned value \hat{V} (black curve, plotted here against objective time) does not drop immediately to zero after reward time *T* (dashed vertical line). This gradual

decrease allows its derivative \hat{V} to be negative over a nonzero domain of time, which in turn allows the update rule for η in Eq. 6 to take negative values over that same domain. It follows that η increases if reward is delivered roughly before T and decreases if reward is delayed past T (gray curve).

and Poole 2012), and emotion (Droit-Volet and Meck 2007), which may serve adaptive functions (Harrington et al. 2011; Matthews and Meck 2014; Meck 2003). To characterize these types of modulation, researchers have predominantly relied on either the peak-interval procedure (Roberts 1981) or the bisection task (Church and Deluty 1977). The peak-interval procedure is a reproduction task: Here, the animal is trained to respond maximally at a criterion duration to receive a reward. Occasionally, much longer, unreinforced probe trials are presented, and the peak response, a measure of the animal's estimate of the time of reward delivery, is measured against the desired manipulation (e.g., administration of DA agonist). In this procedure, acutely increasing the speed of the pacemaker will result in underproduction of a timed interval. The bisection task, on the other hand, is an estimation task: Here, the animal is trained to respond differently to stimuli of either short or long duration, such as pressing a left lever when presented with a short interval and a right lever when presented with a long interval. In probe trials, unreinforced intervals of intermediate length are presented, and the animal must respond with either the "short" response or the "long" response. A psychometric function is fit against each of the control condition and the desired manipulation, and the two functions are then compared. A faster pacemaker will cause interval durations to be perceived as longer than they are, thus resulting in overestimation of intervals. Hence, acutely increasing the pacemaker rate will result in underproduction and overestimation. Similarly, acutely decreasing the pacemaker rate will lead to overproduction and underestimation.

Perhaps the main attraction of a bidirectional learning rule on η is that it allows our model to explain the seemingly conflicting effects of DA, and by extension unexpected rewards, on the pacemaker rate. Indeed, as discussed below, behavioral and pharmacological studies have found that higher reward rates and DA agonists result in behaviors consistent with a faster pacemaker, whereas optogenetic activation of midbrain DA neurons displays the opposite pattern. We show here that these results can be reconciled under the model behaviors derived above.

First, our model is consistent with results from the behavioral literature. In Morgan et al. (1993), pigeons were trained on a bisection task in which they had to discriminate between a 10-s and a 20-s duration and were then exposed to high or low rates of freely delivered reinforcers. When returned to the discrimination task, those returning from a richer context judged durations to be longer than those returning from a poorer context, consistent with a faster and slower pacemaker, respectively (Fig. 5B). Previously, Killeen and Fetterman (1988) had found that the speed of the pacemaker is directly proportional to the rate of reinforcement (Fig. 5A; see also MacEwen and Killeen 1991), similarly predicted by our model. These results are also consistent with studies of electrical brain stimulation of the medial forebrain bundle in mouse, an area that is involved in pleasure sensation and that supports selfstimulation (Olds 1958). Stimulation of this area resulted in a shift of timing behavior that is similarly consistent with a faster pacemaker (Meck 2014).

Second, the model is consistent with results from electrophysiology. As mentioned above, Mello et al. (2015) have shown that in a timekeeping task the response profiles of putative medium spiny neurons (MSNs) representing time cells undergo significant rescaling to reflect the temporal properties of the task at hand (Fig. 6). By construction, this is captured by our model: Here, DA modulates feature rescaling against objective time by exactly this mechanism, and this effect is unexplained by the existing TD and microstimulus models.

Third, the model is consistent with results from pharmacology. Pharmacological manipulations have shown that in a bisection task DA agonists such as methamphetamine predominantly shift the psychometric function to the left (or globally increase the probability of responding "long"), consistent with making the clock go faster, whereas DA antagonists like haloperidol shift it to the right, consistent with making it go slower (Maricq and Church 1983). Lake and Meck (2013) have

Fig. 5. Results from behavioral experiments compared with model behavior. A: Killeen and Fetterman (1988) have found that the speed of the pacemaker is directly proportional to the rate of reinforcement. Reprinted from Killeen and Fetterman (1988). B: Morgan et al. (1993) have found that when exposed to high or low rates of reinforcement and returned to a baseline condition, pigeons' behaviors were consistent with a faster or slower pacemaker, respectively. Reprinted from Morgan et al. (1993) with permission from Elsevier. C and D: model behavior, recapitulating observations in A and B, respectively. See METHODS for simulation details.



Downloaded from www.physiology.org/journal/jn at Harvard Univ (206.253.207.235) on May 1, 2019.

ADAPTING THE FLOW OF TIME WITH DOPAMINE



Fig. 6. Results from electrophysiology compared with model behavior. A: in Mello et al. (2015), electrophysiological recordings in rat striatum during a timekeeping task identified medium spiny neurons that fired sequentially during the delay period and whose response profiles rescaled to reflect the timed duration [fixed interval (FI); note scaling of *x*-axis] but maintained their relative ordering. In addition, by visual inspection, the gradual increase in response profile widths across cells within each trial seems similar across different task durations, as would be suggested by the scalar property. Reprinted from Mello et al. (2015) with permission from Elsevier. *B*: our model recapitulates both phenomena. See METHODS for simulation details.

demonstrated this effect on timekeeping with a peak-interval procedure, which our model recapitulates in Fig. 7.

Finally, our model is consistent with results from optogenetics. In Soares et al. (2016), successive optogenetic stimulation occurred over the duration of the entire trial, which corresponds in our model to sequentially updating η with every new traversed state in that trial by the increment $\Delta \eta$, determined by *Eq.* 6 and the gray curve in Fig. 4. Hence we may approximate the total change in η over the course of an entire trial as

$$\Delta \eta = \sum_{\tau} \alpha_{\eta} \frac{\tau}{\eta} \dot{V}_{\tau} \delta_{\tau} \tag{7}$$

Importantly, the stimulation protocol covered both the early positive limb and the late negative limb of the gray curve in Fig. 4. Under our model, optogenetic activation ($\delta_{\tau} > 0$) therefore resulted in updates of η that first increased it (early stimulation, before the peak of the estimated value function) and then decreased it (late stimulation, after the peak). It is straightforward to show that for optogenetic activation the quantity in Eq. 7 is negative (see appendix for derivation). This can also be visually ascertained from Fig. 4, by noting that the negative limb of the gray curve is larger in area than the positive limb. Therefore, the net effect of this trial-long activation is to slow down the pacemaker. Similarly, trial-long inhibition ($\delta_{\tau} < 0$) will increase η , and the pacemaker will speed up (Fig. 8). Note here that, for convenience, we have taken the effects of optogenetic activation and inhibition to be equal and opposite ($\delta_{\tau} = +1$ and -1, respectively), but this need not be the case. In fact, the experimental data show a rather small effect for inhibition. In principle, this asymmetry may follow from differences in the stimulation protocol, from the asymmetric relationship of DA levels with positive and negative RPEs (Bayer et al. 2007), from the nonlinearity of behavior with equal and opposite changes in value (Luce 1959; Shepard 1958), or from other causes.

In Toda et al. (2017), mice were trained on a peak-interval licking schedule in which a drinking needle delivered sucrose solution every 10 s. Licking behavior was stereotyped, and peak licking occurred at approximately the expected reward time. The authors showed that although start and end times for licking tended to vary with satiety, the peak time was robust to such variations and continued to reflect the expected reward time. The authors then optogenetically stimulated the nigrotectal pathway, an output pathway of the BG, for a 1-s duration that occurred at different times during the trial. Interestingly, stimulation for 1 s immediately before or immediately after reward resulted in later peak responding on the subsequent trial, whereas stimulation ending 1 s before the reward led to earlier peak responding on the subsequent trial.

As illustrated in Fig. 9, this asymmetry follows from our model. Because the sign of \hat{V}_{τ} changes at the peak of the estimated value function, which itself occurs before the true time of reward (Figs. 3 and 4), it follows from Eq. 6 that RPEs that occur after, or immediately before, reward time will have an opposite effect on the speed of the clock compared with those occurring sufficiently before reward time. Importantly, how the output of the BG, reflecting estimated value (Ratcliff and Frank 2012), is transformed and combined into the DA signal from midbrain has been ill defined, although various hints exist (e.g., Barter et al. 2015; Eshel et al. 2015). We assume here that increases in this output serve to increase $\delta_{\tau} = \gamma \hat{V}_{\tau+1} - \hat{V}_{\tau}$. This means that stimulation after, or immediately

1753



Fig. 7. Results from pharmacology compared with model behavior. *A*: in Lake and Meck (2013), subjects reproduced previously learned 7-s and 17-s intervals after administration of the dopamine (DA) agonist amphetamine (AMP), the DA antagonist haloperidol (HAL), or placebo. In the majority of subjects, amphetamine reduced response time, whereas haloperidol delayed it. Reprinted from Lake and Meck (2013) with permission from Elsevier. *B*: our model recapitulates these effects. See METHODS for simulation details.

before, reward time will decrease η and lead to later responding on the next trial. On the other hand, stimulation that is early enough compared with reward time will increase η and lead to earlier responding.

DISCUSSION

The RPE hypothesis of DA is a compelling one, with a rich history of theoretical and experimental support (Eshel et al. 2015; Glimcher 2011; Montague et al. 1996; Niv and Schoenbaum 2008; Schultz 2007; Schultz et al. 1997; Steinberg et al. 2013). In this article, we consider one effect of DA that at first glance does not seem to relate to RPEs: the modulation of timekeeping estimates by DA.

We have shown in this work that, beyond updating the weights w_d of relevant features, the RPE interpretation may be generalizable to learning about other relevant parameters, including, for temporal tasks, the pacemaker rate. Given experimental evidence that DA modulates the pacemaker rate, and inspired by the recent discovery of scalable "time cells," we have shown that the same RPE signal, when extended to the pacemaker rate η , reconciles a wide array of experimental observations, including some that may otherwise seem contradictory. Concretely, we derived from first principles a biologically plausible learning rule on η that facilitates either speed-

ing or slowing the pacemaker, depending on the timing of the DA signal: When reward is presented earlier than expected, the function of η is to increase, thus compressing the estimated value function against objective time, which allows it to better match the true value. On the other hand, when reward is presented late, η must decrease to allow the estimated value to expand appropriately.

A learnable η , and hence a scalable feature set, allows similar numbers of features to participate in the timed interval independent of its length, thus maintaining time cells in the dynamic range of the relevant task. This carries implications for precision, whereby temporal estimates of shorter intervals can be resolved with higher granularity, and may account for the scalar property (Gibbon 1977). In fact, theoretical and experimental work has implicated DA in various measures of precision (Friston et al. 2012; Kroener et al. 2009; Manohar et al. 2015). Second, rescaling the features when the rate of reinforcement changes allows value functions to be rapidly updated without a need for relearning each weight w_d individually. This allows for fewer update steps to occur before convergence (as only a single parameter must now be updated), which paves the way for potentially faster convergence. In addition, we have limited our discussion to interval timing, but this rescaling may be extended to putative downstream representations, such as action velocity. For instance, Yin (2014) has proposed that DA may function as gain control on velocity reference signals, which may explain the bradykinesia of Parkinson's disease and motor symptoms of Tourette syndrome. Indeed, under our formulation, a compression of time cells with high tonic DA will lead to a faster execution of learned action sequences, effectively implementing a type of gain control on executed velocity. Similarly, rescaling time cells to fit the task-specific dynamic range may carry implications for (upstream) cognitive control and attentional modulation, if these inputs exert their influence by operating at the level of the time cells themselves.

The key contribution of the present work-a bidirectional learning rule on the pacemaker rate that critically depends on the timing of DA-is conceptually in line with a number of observations from the broader BG literature. It has been well established that the modification of cortico-striatal weights occurs via a bidirectional, spike-timing-dependent plasticity mechanism (Song et al. 2000) that depends critically on dopaminergic modulation (Pawlak and Kerr 2008; Shen et al. 2008). Emerging experimental evidence furthermore supports the existence of bidirectional mechanisms for learning in striatum that depend on the timing of stimulation. Notably, Yttri and Dudman (2016) have found that optogenetically stimulating MSNs during an approach task can either increase or decrease the subsequent velocity of approach, depending on when during the task the stimulation occurs: Stimulation of D1 MSNs during the fastest phase of the animal's motion in the task induced increases in velocity on both current and subsequent trials, whereas stimulation during the slowest phase of motion induced decreases in velocity. D2 MSNs, which are thought to act antagonistically to D1 MSNs in motor control (Albin et al. 1989; DeLong 1990; Kravitz et al. 2012; Smith et al. 1998), displayed the opposite effect (an increase in velocity following stimulation during the slowest phase and a decrease following stimulation during the fastest phase). For time-



Fig. 8. Results from optogenetic stimulation of midbrain dopamine neurons compared with model behavior. A and B: in Soares et al. (2016), mice were trained on a temporal discrimination task in which they had to judge intervals as either shorter or longer than 1.5 s, and psychometric functions were fit to the data (black curves in both panels). A: under optogenetic activation spanning the entire trial, the psychometric function shifted to the right (dark gray curve), consistent with a slower pacemaker. Insets show the average difference between the probability of selecting the long choice during activation trials vs. control trials per animal (top left) or per stimulus (bottom right). B: under optogenetic inhibition, the psychometric function shifted to the left (light gray curve), consistent with a faster pacemaker. Insets: same as in A, but for inhibition. A and B from Soares et al. Science 354: 1273-1277, 2016. Reprinted with permission from AAAS. C and D: our model recapitulates these effects. See METHODS for simulation details.

keeping in particular, it is the D2 receptor that seems to mediate the effects of DA. For instance, the discussed pharmacological effects on timekeeping relied on haloperidol (Fig. 7), which is a D2 antagonist, whereas these effects were not found with SCH-23390, a D1 antagonist (Drew et al. 2003). Transient overexpression of D2 in striatum has similarly been found to impair timing behaviors (Drew et al. 2007), although the understanding of the roles of D1 and D2 in interval timing is far from complete (see Coull et al. 2011 for a review).

Throughout this article, we have made a number of simplifying assumptions. First, for the optogenetic data, Soares et al. (2016) only observed the effects of dopaminergic modulation of temporal estimates in the currently evaluated trial, rather than in future trials in which modulation of synaptic weights is more likely to play a dominant role. The TD framework is untroubled by this limitation, as weights are not restricted to trial-level learning; rather, learning occurs across states, even within a single trial. Furthermore, as discussed above, in examining the effects of MSN stimulation on movement control, Yttri and Dudman (2016) have found that the timedependent bidirectional effects on the current trial extended to future trials as well, presumably through plasticity-mediated changes. Second, for a canonical timekeeping model, our



Fig. 9. Results from optogenetic stimulation of basal ganglia output compared with model behavior. A, top: in Toda et al. (2017), mice were trained on a peak-interval licking task, and peak licking robustly reflected reward time. The nigrotectal pathway was optogenetically stimulated immediately after reward (*left*), immediately before reward (*center*), or 1 s before reward (*right*). Bottom: stimulation resulted in a shift in peak licking on the subsequent trial. The peak time occurred later when stimulation was delivered immediately after or immediately before reward (*left* and *center*), and it occurred earlier when stimulation was completed 1 s before reward (*right*). *P < 0.05. Reprinted from Toda et al. (2017) with permission from Elsevier. B: our model recapitulates these effects. See METHODS for simulation details.

J Neurophysiol • doi:10.1152/jn.00817.2018 • www.jn.org

framework afforded us the convenience of examining the modulation of timekeeping with a single parameter. Alternative models, described below, can replace this framework with no effect on the bidirectionality of our update rule, which is determined by performing gradient ascent on \hat{V}_{τ} independent of implementation. Third, for our canonical RL model, we have restricted learning to the case in which a nonzero value function has already been established and a change in the temporal dynamics of a simple reward distribution causes a subsequent initiation of learning. Hence, we have assumed that η is learned at a much faster rate than w_d . A series of behavioral papers have indeed shown rapid, nearly one-shot or two-shot learning of temporal duration in rats (Davis et al. 1989; Higa 1997; Meck et al. 1984), mice (Balci et al. 2008, with reanalysis in Simen et al. 2011, Supplemental Fig. 10), pigeons (Wynne and Staddon 1988), and humans (Simen et al. 2011). Recent theoretical work (Simen et al. 2011), discussed below, has taken these findings into account and developed learning rules that are particularly effective at rapid, single-exposure learning, compared with the TD models we have assumed, where learning is more gradual.

Finally, this work inherits the limitations of the existing literature on DA and interval timing. For instance, the present findings of DA's role in timekeeping do not preclude the possibility of unanticipated off-target effects. Similarly, beyond influencing the internal clock, DA may affect motivation (Balci 2014; Galtress et al. 2012), attention (Nieoullon 2002), and the role of context (Gu et al. 2015; Malapani et al. 1998; Shi et al. 2013) and may asymmetrically influence the encoding and decoding of temporal estimates in memory (Malapani et al. 2002). Although these are not novel concepts and time-keeping experiments have attempted to control for such factors (e.g., Lake and Meck 2013; Soares et al. 2016), the potential for confounds remains.

Experimental predictions. Our model makes a number of novel experimental predictions. First, the model makes a strong claim on the bidirectional rule governing the update of η , as written in Eq. 6 and illustrated in Fig. 4. This claim can be experimentally verified with the paradigm of Soares et al. (2016): Correlating changes in pacemaker rate with DA levels preceding choice time by some increment Δt , for each increment independently, will recapitulate the gray curve in Fig. 4. In addition, direct optogenetic manipulation of DA levels over different intervals within a single trial will demonstrate this same relationship between DA levels and pacemaker rate, but with causality.

Second, as a result of feature rescaling, our model makes a strong claim about the conditioned stimulus (CS) response in classical conditioning tasks with fixed delay and in particular how it should change under different environmental or pharmacological conditions. Consider then the case in which animals are both trained and tested on timing tasks under either high or low tonic DA conditions. We have shown in RESULTS that high tonic DA leads to feature compression under our model. It follows that animals experiencing high-DA conditions will experience more states during the same delay period. This means that value at the CS will be of smaller magnitude than that of animals trained and tested under low-DA conditions (Fig. 10*A*). Hence, the phasic DA CS response will also be smaller under higher-DA conditions (Fig. 10*B*). This experiment can similarly be performed by varying reward rate



Fig. 10. Simulated effect of tonic dopamine (DA) on value and reward prediction error (RPE) against objective time. A: in our model, higher-DA conditions lead to a faster pacemaker and more compressed features, which serve as states. By Eq. 4, this leads to a steeper slope in the value function when measured against objective time. B: DA response, computed by Eq. 4, and based on the corresponding value functions in A. Left and right dashed lines denote conditioned stimulus and reward time, respectively.

environments rather than DA levels, with similar predicted results.

Third, our model asserts that the pacemaker rate and external reward rate are state functions. In other words, given a current average reward rate, η will converge to the same value regardless of reward history. As described in RESULTS, Morgan et al. (1993) studied the immediate effect of reward history on pigeons in a discrimination task and found that pigeons returning from a highly rewarding context initially judged duration to be longer than those returning from a poorer context, and these judgments in both groups slowly adjusted to the new environment. Our additional prediction, however, is that after overcoming this immediate effect the pacemaker rate in both conditions will converge to the same value (see Fig. 5*B*).

Fourth, our model makes a claim about the interaction between tonic and phasic DA levels, namely, that tonic DA functions to bias the pacemaker toward speeding up, whereas phasic DA functions to modulate the pacemaker in either direction depending on its timing. This interaction can be tested with combinations of experimental manipulations, for instance, by varying either the average reward rate or the type of optogenetic stimulation (activation vs. inhibition) against a background of either high or low tonic DA levels. The behavioral output of a categorization task similar to that of Soares et al. (2016) can address this prediction.

Relationship with other models. As mentioned above, our key result, that a DA-dependent bidirectional rule controls time cell response scaling, can be incorporated into a large

class of timing models, including pacemaker-accumulator (PA) models, sequential-state models, and striatal beat-frequency models.

In the influential PA model (Gibbon et al. 1997; Treisman 1963; Zakay and Block 1997), a pacemaker, representing the internal clock, emits pulses with approximately constant periodicity. When an interval is to be timed, an accumulator collects the pulses emitted during that interval, and the total number of accumulated pulses is subsequently stored in memory. To reproduce the stored interval, accumulation begins afresh and continues until the accumulated quantity closely matches the quantity stored in memory. This model, whose terminology we have borrowed, maps very naturally onto our framework: Here, the second layer of our neural circuit, encoding subjective time τ , represents the pacemaker, and the time cells, taken as an ordered set, represent the accumulator. The modulation of the pacemaker rate occurs by changing η .

Despite its abstract nature, perhaps the main attraction of the PA model is its intuitive implementation of timekeeping and its biases. For instance, acutely speeding the internal clock results in underproduction and overestimation of intervals. Under the PA model, this is intuitive: During reproduction of a previously stored interval, acutely increasing the speed of the pacemaker causes the accumulator to collect the desired number of pulses within a shorter time frame (underproduction). Alternatively, during a bisection task an acutely sped-up pacemaker will cause the accumulator to collect more pulses by the end of the presented interval than it otherwise would have, thus creating a bias toward selecting the "long" response (overestimation). Similarly, acutely slowing the pacemaker will lead to overproduction and underestimation.

Sequential-state models are similar to PA models but assume a population of cells that are connected in series (rather than in parallel) and thus sequentially activate each other with a time delay. These cells closely resemble the time cells in our model, and their activations serve to encode time (Buonomano and Merzenich 1995; Killeen and Fetterman 1988). For instance, in the behavioral theory of timing (BeT) (Killeen and Fetterman 1988), the cells correspond to behavioral states and the rate of transition across states is taken to be proportional to the reward rate. LeT (Machado 1997) is a descendant of BeT that further views these states as a basis set whose weighted sum constitutes the output layer. Although not a normatively justified RL model, LeT succeeds in proposing associative rules to govern these weights, which allow rewards to be correctly associated with early or late states. This is akin to the bottom two layers in our model and represents an early effort to use principles of RL in interval timing.

Finally, striatal beat-frequency models assume oscillatory neurons in cortex that fire and converge on MSNs in striatum. Timing information is stored in the cortico-striatal weights, gated by DA-dependent long-term potentiation and depression. Subsequent comparisons of time durations with the stored memory are then achieved by comparing the pattern of activation of the oscillatory neurons with the memory trace (Matell and Meck 2000, 2004). Although the architectures of the three presented models differ, it remains that a gradient ascent rule on a parameter controlling the rate of transitions between states, seeking to minimize the difference between the true value V_{τ} and estimated value \hat{V}_{τ} , will qualitatively recapitulate the bidirectionality derived in Eq. 6.

With respect to the conflicting experimental evidence regarding the role of DA in timekeeping, we are not aware of any computational work that has sought to reconcile these observations. However, the idea of parameterized rescaling, interpreted broadly, is not novel in the interval timing literature, and whether DA may be related to these models is an open question. For instance, modulating drift rates for the purpose of learning to time is a key component of drift-diffusion models of interval timing (Luzardo et al. 2013; Rivest and Bengio 2011; Simen et al. 2011), and recent work has related these models with theories of classical conditioning (Luzardo et al. 2017). In the time-adaptive opponent Poisson drift-diffusion model (TopDDM) (Balc1 and Simen 2016; Rivest and Bengio 2011; Simen et al. 2011), the ramping rate of neural integrators (or the drift) to a fixed threshold is controlled by the proportion of its active inputs, whose activations are in turn controlled by learning rules on their own input weights (see Komura et al. 2001 for a potential neural correlate of the integrator in sensory thalamus). A virtue of these update rules is that complete adaptation to a new duration can occur after a single exposure: If the timed duration ends before the ramp reaches the fixed threshold (late timing), then a "late learning" rule increases the input weights to the strength required for accurate timing of the new duration. On the other hand, if the threshold is reached before the end of the duration (early timing), a separate "early learning" rule begins to decrease the weights continuously until the timed duration ends. This occurs in such a way that the new drift will allow for accurate timing of the new duration. As discussed above, such learning rules, for either increasing or decreasing drift, are particularly adept at capturing rapid and one-shot learning.

Finally, there is a substantial literature within the study of interval timing that examines the effects of attentional control on timekeeping (Fortin 2003). It is conceivable that the different effects of DA may be explained by invoking differential attentional modulation, for instance, by reallocating the limited resource of attention either away from or toward the timekeeping system, depending on the experimental paradigm (Meck 1984). It is furthermore possible that attentional control works in tandem with the update mechanism proposed here, for instance, by modulating the compression t' of objective time t or by other mechanisms. Our aim is not to refute these possibilities but instead to show how a single RPE interpretation can account for a wide range of seemingly conflicting results.

APPENDIX

Alternative implementations of the pacemaker rate. Rather than introducing a separate layer representing t' in our circuit, it is possible to directly modulate the weights between τ and $x_{d,\tau}$ by η_{d} . This is a feasible alternative; however, without restricting the parameter space or introducing additional assumptions, such a neural circuit may result in crossovers of the time cell responses. For instance, if a large reward is delivered at $\tau = 3$, the update of η_3 may be enough to cause $x_{3,\tau}$ to respond earlier than $x_{2,\tau}$ on the next trial. By restricting learning to a single η , our proposed circuit protects against this possibility. We can show this by rewriting $x_{d,\tau}$ as

ADAPTING THE FLOW OF TIME WITH DOPAMINE

$$x_{d,\tau} = \exp\left[-\frac{1}{2}\left(\frac{\eta t' - \mu_d}{\sigma}\right)^2\right]$$
$$= \exp\left[-\frac{1}{2}\left(\frac{t' - \frac{\mu_d}{\eta}}{\frac{\sigma}{\eta}}\right)^2\right]$$

Hence, samples of x_d are drawn from $\mathcal{N}\left[\frac{\mu_d}{\eta}, \left(\frac{\sigma}{\eta}\right)^2\right]$ up to a constant, and the coefficient of variation is $CV = \frac{\sigma}{\eta}\frac{\eta}{\mu_d} = \frac{\sigma}{\mu_d}$, which

is independent of η . In addition, the means for all features x_d are $\frac{\mu_d}{\eta}$. Therefore the ordering of μ_d is maintained when scaling by η , and no crossovers occur.

Sequential activation of DA neurons over the course of an entire trial results in a net negative RPE. Let us analytically derive this result. We begin by writing

$$\Delta \eta = \sum_{\tau} \alpha_{\eta} \frac{\tau}{\eta} \dot{V}_{\tau} \delta_{\tau} \tag{A1}$$

over the duration of an entire trial. For a better intuition of the result we will derive below, let us extend our analysis to the continuous domain. The relationship in Eq. A1 thus becomes

$$\Delta \eta = \int_{\tau} \alpha_{\eta} \frac{\tau}{\eta} \dot{V}_{\tau} \delta_{\tau} d\tau \tag{A2}$$

Because the estimated value \hat{V}_τ decays to zero after reward time $\mathcal{T},$ it follows that

$$\int_{\tau} \dot{\hat{V}}_{\tau} d\tau = 0 \tag{A3}$$

Here, \hat{V}_{τ} , like $\Delta \eta$, has a positive limb, or a domain over which \hat{V}_{τ} is >0, followed by a negative limb. Additionally, by Eq. A3, the integral over the positive limb is equal and opposite to that over the negative limb.

Since $\frac{\tau}{\eta} = t'$ is monotonically increasing, it follows that

$$\int_{\tau} \frac{\tau}{\eta} \dot{\hat{V}}_{\tau} d\tau < 0$$

Therefore, with optogenetic activation ($\delta_{\tau} > 0$), the integral in Eq. A2 is negative, so that η will decrease and the pacemaker will slow down. With a similar analysis for optogenetic inhibition, η will increase and the pacemaker will speed up.

ACKNOWLEDGMENTS

The authors thank John Assad and Allison Hamilos for helpful discussions.

GRANTS

The project described was supported by National Institute of General Medical Sciences Grant T32 GM-007753 (J. G. Mikhael) and National Institute of Mental Health Grant CRCNS 1R01 MH-109177 (S. J. Gershman).

DISCLAIMERS

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

J.G.M. and S.J.G. conceived and designed research; J.G.M. analyzed data; J.G.M. prepared figures; J.G.M. drafted manuscript; J.G.M. and S.J.G. edited and revised manuscript; J.G.M. and S.J.G. approved final version of manuscript.

ENDNOTE

At the request of the authors, readers are herein alerted to the fact that additional materials related to this manuscript may be found at GitHub. These materials are not a part of this manuscript, and have not undergone peer review by the American Physiological Society (APS). APS and the journal editors take no responsibility for these materials, for the website address, or for any links to or from it.

REFERENCES

- Albin RL, Young AB, Penney JB. The functional anatomy of basal ganglia disorders. *Trends Neurosci* 12: 366–375, 1989. doi:10.1016/0166-2236(89)90074-X.
- Allman MJ, Teki S, Griffiths TD, Meck WH. Properties of the internal clock: first- and second-order principles of subjective time. Annu Rev Psychol 65: 743–771, 2014. doi:10.1146/annurev-psych-010213-115117.
- Artieda J, Pastor MA, Lacruz F, Obeso JA. Temporal discrimination is abnormal in Parkinson's disease. *Brain* 115: 199–210, 1992. doi:10.1093/ brain/115.1.199.
- Arushanyan E, Baida O, Mastyagin S, Popova A, Shikina I. Influence of caffeine on the subjective perception of time by healthy subjects in dependence on various factors. *Hum Physiol* 29: 433–436, 2003. doi:10.1023/ A:1024973305920.
- **Balcı F.** Interval timing, dopamine, and motivation. *Timing Time Percept* 2: 379–410, 2014. doi:10.1163/22134468-00002035.
- Balci F, Papachristos EB, Gallistel CR, Brunner D, Gibson J, Shumyatsky GP. Interval timing in genetically modified mice: a simple paradigm. *Genes Brain Behav* 7: 373–384, 2008. doi:10.1111/j.1601-183X.2007.00348.x.
- Balci F, Simen P. A decision model of timing. *Curr Opin Behav Sci* 8: 94–101, 2016. doi:10.1016/j.cobeha.2016.02.002.
- Barter JW, Li S, Lu D, Bartholomew RA, Rossi MA, Shoemaker CT, Salas-Meza D, Gaidis E, Yin HH. Beyond reward prediction errors: the role of dopamine in movement kinematics. *Front Integr Neurosci* 9: 39, 2015. doi:10.3389/fnint.2015.00039.
- Bayer HM, Lau B, Glimcher PW. Statistics of midbrain dopamine neuron spike trains in the awake primate. J Neurophysiol 98: 1428–1439, 2007. doi:10.1152/jn.01140.2006.
- Buhusi CV, Meck WH. What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci* 6: 755–765, 2005. doi:10.1038/ nrn1764.
- Buonomano DV, Merzenich MM. Temporal information transformed into a spatial code by a neural network with realistic properties. *Science* 267: 1028–1030, 1995. doi:10.1126/science.7863330.
- Cheng RK, Ali YM, Meck WH. Ketamine "unlocks" the reduced clock-speed effects of cocaine following extended training: evidence for dopamine– glutamate interactions in timing and time perception. *Neurobiol Learn Mem* 88: 149–159, 2007. doi:10.1016/j.nlm.2007.04.005.
- Church RM, Deluty MZ. Bisection of temporal intervals. J Exp Psychol Anim Behav Process 3: 216–228, 1977. doi:10.1037/0097-7403.3.3.216.
- Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J Neurosci 29: 1538–1543, 2009. doi:10.1523/JNEUROSCI.4467-08.2009.
- Coull JT, Cheng RK, Meck WH. Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology* 36: 3–25, 2011. doi:10. 1038/npp.2010.113.
- Coull JT, Vidal F, Nazarian B, Macar F. Functional anatomy of the attentional modulation of time estimation. *Science* 303: 1506–1508, 2004. doi:10.1126/science.1091573.
- Davis M, Schlesinger LS, Sorenson CA. Temporal specificity of fear conditioning: effects of different conditioned stimulus-unconditioned stimulus

intervals on the fear-potentiated startle effect. J Exp Psychol Anim Behav Process 15: 295–310, 1989. doi:10.1037/0097-7403.15.4.295.

- DeLong MR. Primate models of movement disorders of basal ganglia origin. Trends Neurosci 13: 281–285, 1990. doi:10.1016/0166-2236(90)90110-V.
- Drew MR, Fairhurst S, Malapani C, Horvitz JC, Balsam PD. Effects of dopamine antagonists on the timing of two intervals. *Pharmacol Biochem Behav* 75: 9–15, 2003. doi:10.1016/S0091-3057(03)00036-4.
- Drew MR, Simpson EH, Kellendonk C, Herzberg WG, Lipatova O, Fairhurst S, Kandel ER, Malapani C, Balsam PD. Transient overexpression of striatal D2 receptors impairs operant motivation and interval timing. *J Neurosci* 27: 7731–7739, 2007. doi:10.1523/JNEUROSCI.1736-07.2007.
- **Droit-Volet S, Meck WH.** How emotions colour our perception of time. *Trends Cogn Sci* 11: 504–513, 2007. doi:10.1016/j.tics.2007.09.008.
- Elvevåg B, McCormack T, Gilbert A, Brown GD, Weinberger DR, Goldberg TE. Duration judgements in patients with schizophrenia. *Psychol Med* 33: 1249–1261, 2003. doi:10.1017/S0033291703008122.
- Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N. Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* 525: 243–246, 2015. [Erratum in *Nature* 527: 398, 2015.] doi:10.1038/ nature14855.
- Fortin, C. Attentional time-sharing in interval timing. In: *Functional and Neural Mechanisms of Interval Timing*, edited by Meck WH. Boca Raton, FL: CRC, 2003, p. 235–260.
- Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci USA* 104: 16311–16316, 2007. doi:10.1073/ pnas.0706111104.
- Frank MJ, Seeberger LC, O'reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* 306: 1940–1943, 2004. doi:10.1126/science.1102941.
- Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan KE, Bestmann S. Dopamine, affordance and active inference. *PLoS Comput Biol* 8: e1002327, 2012. doi:10.1371/journal.pcbi. 1002327.
- Gable PA, Poole BD. Time flies when you're having approach-motivated fun: effects of motivational intensity on time perception. *Psychol Sci* 23: 879–886, 2012. doi:10.1177/0956797611435817.
- Galtress T, Marshall AT, Kirkpatrick K. Motivation and timing: clues for modeling the reward system. *Behav Processes* 90: 142–153, 2012. doi:10. 1016/j.beproc.2012.02.014.
- Gershman SJ, Moustafa AA, Ludvig EA. Time representation in reinforcement learning models of the basal ganglia. *Front Comput Neurosci* 7: 194, 2014. doi:10.3389/fncom.2013.00194.
- Gibbon J. Scalar expectancy theory and Weber's law in animal timing. *Psychol Rev* 84: 279–325, 1977. doi:10.1037/0033-295X.84.3.279.
- Gibbon J, Malapani C, Dale CL, Gallistel C. Toward a neurobiology of temporal cognition: advances and challenges. *Curr Opin Neurobiol* 7: 170–184, 1997. doi:10.1016/S0959-4388(97)80005-0.
- Glimcher PW. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc Natl Acad Sci USA* 108, *Suppl* 3: 15647–15654, 2011. [Erratum in *Proc Natl Acad Sci USA* 108: 17568–17569, 2011.] doi:10.1073/pnas.1014269108.
- Gu BM, Jurkowski AJ, Lake JI, Malapani C, Meck WH. Bayesian models of interval timing and distortions in temporal memory as a function of Parkinson's Disease and dopamine-related error processing. In: *Time Distortions in Mind: Temporal Processing in Clinical Populations*, edited by Vatakis A, Allman MJ. Boston, MA: Brill, 2015, p. 281–327.
- Harrington DL, Castillo GN, Fong CH, Reed JD. Neural underpinnings of distortions in the experience of time across senses. *Front Integr Neurosci* 5: 32, 2011. doi:10.3389/fnint.2011.00032.
- Higa JJ. Rapid timing of a single transition in interfood interval duration by rats. Anim Learn Behav 25: 177–184, 1997. doi:10.3758/BF03199056.
- Houk JC, Adams JL, Barto AG. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser DG. Cambridge, MA: MIT Press, 1995.
- Jahanshahi M, Jones CR, Zijlmans J, Katzenschlager R, Lee L, Quinn N, Frith CD, Lees AJ. Dopaminergic modulation of striato-frontal connectivity during motor timing in Parkinson's disease. *Brain* 133: 727–745, 2010. doi:10.1093/brain/awq012.
- Killeen PR, Fetterman JG. A behavioral theory of timing. *Psychol Rev* 95: 274–295, 1988. doi:10.1037/0033-295X.95.2.274.
- Klopf AH. A neuronal model of classical conditioning. *Psychobiology (Austin Tex)* 16: 85–125, 1988.

- Kobayashi S, Schultz W. Influence of reward delays on responses of dopamine neurons. J Neurosci 28: 7837–7846, 2008. doi:10.1523/JNEUROSCI. 1600-08.2008.
- Komura Y, Tamura R, Uwano T, Nishijo H, Kaga K, Ono T. Retrospective and prospective coding for predicted reward in the sensory thalamus. *Nature* 412: 546–549, 2001. doi:10.1038/35087595.
- Kosko B. Differential Hebbian learning. In: AIP Conference Proceedings 151 on Neural Networks for Computing. Woodbury, NY: American Institute of Physics, 1986, p. 277–282.
- Kravitz AV, Tye LD, Kreitzer AC. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat Neurosci* 15: 816–818, 2012. doi:10.1038/nn.3100.
- Kroener S, Chandler LJ, Phillips PE, Seamans JK. Dopamine modulates persistent synaptic activity and enhances the signal-to-noise ratio in the prefrontal cortex. *PLoS One* 4: e6507, 2009. doi:10.1371/journal.pone. 0006507.
- Lake JI, Meck WH. Differential effects of amphetamine and haloperidol on temporal reproduction: dopaminergic regulation of attention and clock speed. *Neuropsychologia* 51: 284–292, 2013. doi:10.1016/j.neuropsychologia.2012. 09.014.
- Luce RD. Individual Choice Behavior: A Theoretical Analysis. New York: Wiley, 1959.
- Ludvig E, Sutton RS, Kehoe EJ. Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. *Neural Comput* 20: 3034–3054, 2008. doi:10.1162/neco.2008.11-07-654.
- Ludvig EA, Sutton RS, Kehoe EJ. Evaluating the TD model of classical conditioning. *Learn Behav* 40: 305–319, 2012. doi:10.3758/s13420-012-0082-6.
- Luzardo A, Alonso E, Mondragón E. A Rescorla-Wagner drift-diffusion model of conditioning and timing. *PLoS Comput Biol* 13: e1005796, 2017. doi:10.1371/journal.pcbi.1005796.
- Luzardo A, Ludvig EA, Rivest F. An adaptive drift-diffusion model of interval timing dynamics. *Behav Processes* 95: 90–99, 2013. doi:10.1016/ j.beproc.2013.02.003.
- MacEwen D, Killeen P. The effects of rate and amount of reinforcement on the speed of the pacemaker in pigeons' timing behavior. *Learn Behav* 19: 164–170, 1991. doi:10.3758/BF03197872.
- Machado A. Learning the temporal dynamics of behavior. *Psychol Rev* 104: 241–265, 1997. doi:10.1037/0033-295X.104.2.241.
- Malapani C, Deweer B, Gibbon J. Separating storage from retrieval dysfunction of temporal memory in Parkinson's disease. J Cogn Neurosci 14: 311–322, 2002. doi:10.1162/089892902317236920.
- Malapani C, Rakitin B, Levy R, Meck WH, Deweer B, Dubois B, Gibbon J. Coupled temporal memories in Parkinson's disease: a dopamine-related dysfunction. J Cogn Neurosci 10: 316–331, 1998. doi:10.1162/089892998562762.
- Manohar SG, Chong TT-J, Apps MA, Batla A, Stamelou M, Jarman PR, Bhatia KP, Husain M. Reward pays the cost of noise reduction in motor and cognitive control. *Curr Biol* 25: 1707–1716, 2015. doi:10.1016/j.cub. 2015.05.038.
- Maricq AV, Church RM. The differential effects of haloperidol and methamphetamine on time estimation in the rat. *Psychopharmacology (Berl)* 79: 10–15, 1983. doi:10.1007/BF00433008.
- Maricq AV, Roberts S, Church RM. Methamphetamine and time estimation. *J Exp Psychol Anim Behav Process* 7: 18–30, 1981. doi:10.1037/0097-7403.7.1.18.
- Matell MS, Meck WH. Neuropsychological mechanisms of interval timing behavior. *BioEssays* 22: 94–103, 2000. doi:10.1002/(SICI)1521-1878 (200001)22:1<94:AID-BIES14>3.0.CO;2-E.
- Matell MS, Meck WH. Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Brain Res Cogn Brain Res* 21: 139–170, 2004. doi:10.1016/j.cogbrainres.2004.06.012.
- Matthews WJ, Meck WH. Time perception: the bad news and the good. Wiley Interdiscip Rev Cogn Sci 5: 429–446, 2014. doi:10.1002/wcs.1298.
- Meck WH. Attentional bias between modalities: effect on the internal clock, memory, and decision stages used in animal time discrimination. Ann NY Acad Sci 423: 528–541, 1984. doi:10.1111/j.1749-6632.1984.tb23457.x.
- Meck WH. Affinity for the dopamine D2 receptor predicts neuroleptic potency in decreasing the speed of an internal clock. *Pharmacol Biochem Behav* 25: 1185–1189, 1986. doi:10.1016/0091-3057(86)90109-7.
- Meck WH (Editor). Functional and Neural Mechanisms of Interval Timing. Boca Raton, FL: CRC, 2003.
- Meck WH. Neuroanatomical localization of an internal clock: a functional link between mesolimbic, nigrostriatal, and mesocortical dopaminergic systems. *Brain Res* 1109: 93–107, 2006. doi:10.1016/j.brainres.2006.06.031.

- Meck WH. Internal clock and reward pathways share physiologically similar information-processing stages. In: *Quantitative Analyses of Behavior. Biological Determinants of Reinforcement*, edited by Commons ML, Church RM, Stellar IR, Wagner AR. Hillsdale, NJ: Erlbaum, 2014, p. 121–138.
- Meck WH, Komeily-Zadeh FN, Church RM. Two-step acquisition: modification of an internal clock's criterion. J Exp Psychol Anim Behav Process 10: 297–306, 1984. doi:10.1037/0097-7403.10.3.297.
- Mello GB, Soares S, Paton JJ. A scalable population code for time in the striatum. Curr Biol 25: 1113–1122, 2015. doi:10.1016/j.cub.2015.02.036.
- Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16: 1936–1947, 1996. doi:10.1523/JNEUROSCI.16-05-01936.1996.
- Morgan L, Killeen PR, Fetterman JG. Changing rates of reinforcement perturbs the flow of time. *Behav Processes* 30: 259–271, 1993. doi:10.1016/ 0376-6357(93)90138-H.
- Nieoullon A. Dopamine and the regulation of cognition and attention. *Prog* Neurobiol 67: 53–83, 2002. doi:10.1016/S0301-0082(02)00011-4.
- Niv Y, Schoenbaum G. Dialogues on prediction errors. *Trends Cogn Sci* 12: 265–272, 2008. doi:10.1016/j.tics.2008.03.006.
- **Olds J.** Self-stimulation of the brain; its use to study local effects of hunger, sex, and drugs. *Science* 127: 315–324, 1958. doi:10.1126/science.127.3294. 315.
- Pawlak V, Kerr JN. Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. J Neurosci 28: 2435–2446, 2008. doi:10.1523/JNEUROSCI.4402-07.2008.
- Petter EA, Gershman SJ, Meck WH. Integrating models of interval timing and reinforcement learning. *Trends Cogn Sci* 22: 911–922, 2018. doi:10. 1016/j.tics.2018.08.004.
- Rao RP, Sejnowski TJ. Spike-timing-dependent Hebbian plasticity as temporal difference learning. *Neural Comput* 13: 2221–2237, 2001. doi:10. 1162/089976601750541787.
- Ratcliff R, Frank MJ. Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. *Neural Comput* 24: 1186–1229, 2012. doi:10.1162/NECO_a_00270.
- Reynolds JN, Wickens JR. Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw* 15: 507–521, 2002. doi:10.1016/S0893-6080(02) 00045-X.
- Rivest F, Bengio Y. Adaptive drift-diffusion process to learn time intervals (Preprint). *arXiv* 1103.2382, 2011.
- Roberts PD. Computational consequences of temporally asymmetric learning rules: I. Differential hebbian learning. *J Comput Neurosci* 7: 235–246, 1999. doi:10.1023/A:1008910918445.
- Roberts S. Isolation of an internal clock. J Exp Psychol Anim Behav Process 7: 242–268, 1981. doi:10.1037/0097-7403.7.3.242.
- Rowe KC, Paulsen JS, Langbehn DR, Duff K, Beglinger LJ, Wang C, O'Rourke JJ, Stout JC, Moser DJ. Self-paced timing detects and tracks change in prodromal Huntington disease. *Neuropsychology* 24: 435–442, 2010. doi:10.1037/a0018905.
- Schultz W. Behavioral dopamine signals. *Trends Neurosci* 30: 203–210, 2007. doi:10.1016/j.tins.2007.03.007.
- Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 275: 1593–1599, 1997. doi:10.1126/science.275.5306.1593.
- Shen W, Flajolet M, Greengard P, Surmeier DJ. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321: 848–851, 2008. doi:10.1126/science.1160575.

- Shepard RN. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. J Exp Psychol 55: 509– 523, 1958. doi:10.1037/h0042354.
- Shi Z, Church RM, Meck WH. Bayesian optimization of time perception. *Trends Cogn Sci* 17: 556–564, 2013. doi:10.1016/j.tics.2013.09.009.
- Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P, Dolan RJ. Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. *Brain* 135: 1871–1883, 2012. doi:10.1093/brain/aws083.
- Simen P, Balci F, de Souza L, Cohen JD, Holmes P. A model of interval timing by neural integration. J Neurosci 31: 9238–9253, 2011. doi:10.1523/ JNEUROSCI.3121-10.2011.
- Smith Y, Bevan MD, Shink E, Bolam JP. Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience* 86: 353–387, 1998.
- Smittenaar P, Chase HW, Aarts E, Nusselein B, Bloem BR, Cools R. Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection—learning or performance? *Eur J Neurosci* 35: 1144–1151, 2012. doi:10.1111/j.1460-9568.2012.08043.x.
- Soares S, Atallah BV, Paton JJ. Midbrain dopamine neurons control judgment of time. Science 354: 1273–1277, 2016. doi:10.1126/science.aah5234.
- Song S, Miller KD, Abbott LF. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat Neurosci* 3: 919–926, 2000. doi:10.1038/78829.
- Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 16: 966–973, 2013. doi:10.1038/nn.3413.
- Sutton RS. Learning to predict by the methods of temporal differences. *Mach Learn* 3: 9–44, 1988. doi:10.1007/BF00115009.
- Toda K, Lusk NA, Watson GD, Kim N, Lu D, Li HE, Meck WH, Yin HH. Nigrotectal stimulation stops interval timing in mice. *Curr Biol* 27: 3763–3770.e3, 2017. doi:10.1016/j.cub.2017.11.003.
- **Treisman M.** Temporal discrimination and the indifference interval. Implications for a model of the "internal clock". *Psychol Monogr* 77: 1–31, 1963. doi:10.1037/h0093864.
- Utter AA, Basso MA. The basal ganglia: an overview of circuits and function. *Neurosci Biobehav Rev* 32: 333–342, 2008. doi:10.1016/j.neubiorev.2006. 11.003.
- Wang J, Narain D, Hosseini EA, Jazayeri M. Flexible timing by temporal scaling of cortical responses. *Nat Neurosci* 21: 102–110, 2018. doi:10.1038/ s41593-017-0028-6.
- Ward RD, Kellendonk C, Simpson EH, Lipatova O, Drew MR, Fairhurst S, Kandel ER, Balsam PD. Impaired timing precision produced by striatal D2 receptor overexpression is mediated by cognitive and motivational deficits. *Behav Neurosci* 123: 720–730, 2009. doi:10.1037/a0016503.
- Wynne CD, Staddon JE. Typical delay determines waiting time on periodicfood schedules: static and dynamic tests. J Exp Anal Behav 50: 197–210, 1988. doi:10.1901/jeab.1988.50-197.
- Yin HH. Action, time and the basal ganglia. *Philos Trans R Soc Lond B Biol Sci* 369: 20120473, 2014. doi:10.1098/rstb.2012.0473.
- Yttri EA, Dudman JT. Opponent and bidirectional control of movement velocity in the basal ganglia. *Nature* 533: 402–406, 2016. doi:10.1038/ nature17639.
- Zakay D, Block RA. Temporal cognition. Curr Dir Psychol Sci 6: 12–16, 1997. doi:10.1111/1467-8721.ep11512604.