

RESEARCH ARTICLE

Rational inattention and tonic dopamine

John G. Mikhael^{1,2*}, Lucy Lai¹, Samuel J. Gershman³**1** Program in Neuroscience, Harvard Medical School, Boston, Massachusetts, United States of America,**2** MD-PhD Program, Harvard Medical School, Boston, Massachusetts, United States of America,**3** Department of Psychology and Center for Brain Science, Harvard University, Cambridge, Massachusetts, United States of America* john_mikhael@hms.harvard.edu

Abstract

Slow-timescale (tonic) changes in dopamine (DA) contribute to a wide variety of processes in reinforcement learning, interval timing, and other domains. Furthermore, changes in tonic DA exert distinct effects depending on when they occur (e.g., during learning vs. performance) and what task the subject is performing (e.g., operant vs. classical conditioning). Two influential theories of tonic DA—the average reward theory and the Bayesian theory in which DA controls precision—have each been successful at explaining a subset of empirical findings. But how the same DA signal performs two seemingly distinct functions without creating cross-talk is not well understood. Here we reconcile the two theories under the unifying framework of ‘rational inattention,’ which (1) conceptually links average reward and precision, (2) outlines how DA manipulations affect this relationship, and in so doing, (3) captures new empirical phenomena. In brief, rational inattention asserts that agents can increase their precision in a task (and thus improve their performance) by paying a cognitive cost. Crucially, whether this cost is worth paying depends on average reward availability, reported by DA. The monotonic relationship between average reward and precision means that the DA signal contains the information necessary to retrieve the precision. When this information is needed after the task is performed, as presumed by Bayesian inference, acute manipulations of DA will bias behavior in predictable ways. We show how this framework reconciles a remarkably large collection of experimental findings. In reinforcement learning, the rational inattention framework predicts that learning from positive and negative feedback should be enhanced in high and low DA states, respectively, and that DA should tip the exploration-exploitation balance toward exploitation. In interval timing, this framework predicts that DA should increase the speed of the internal clock and decrease the extent of interference by other temporal stimuli during temporal reproduction (the central tendency effect). Finally, rational inattention makes the new predictions that these effects should be critically dependent on the controllability of rewards, that post-reward delays in intertemporal choice tasks should be underestimated, and that average reward manipulations should affect the speed of the clock—thus capturing empirical findings that are unexplained by either theory alone. Our results suggest that a common computational repertoire may underlie the seemingly heterogeneous roles of DA.

OPEN ACCESS

Citation: Mikhael JG, Lai L, Gershman SJ (2021) Rational inattention and tonic dopamine. *PLoS Comput Biol* 17(3): e1008659. <https://doi.org/10.1371/journal.pcbi.1008659>

Editor: Jean Daunizeau, Brain and Spine Institute (ICM), FRANCE

Received: October 15, 2019

Accepted: December 28, 2020

Published: March 24, 2021

Copyright: © 2021 Mikhael et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Source code for all simulations can be found at www.github.com/jgmikhael/rationalinattention.

Funding: The project described was supported by National Institutes of Health grants T32GM007753 (JGM), T32MH020017 (JGM), and U19 NS113201-01 (SJG), and National Science Foundation Graduate Research Fellowship grant DGE-1745303 (LL). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation. The funders had no role in study

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

The roles of tonic dopamine (DA) have been the subject of much speculation, partly due to the variety of processes it has been implicated in. For instance, tonic DA modulates how we learn new information, but also affects how previously learned information is used. DA affects the speed of our internal timing mechanism, but also modulates the degree to which our temporal estimates are influenced by context. DA improves performance in some tasks, but seems only to affect confidence in others. Are there common principles that govern the role of DA across these domains? In this work, we introduce the concept of ‘rational inattention,’ originally borrowed from economics, to the DA literature. We show how the rational inattention account of DA unites two influential theories that are seemingly at odds: the average reward theory and the Bayesian theory of tonic DA. We then show how this framework reconciles the diverse roles of DA, which cannot be addressed by either theory alone.

Introduction

The functions of dopamine (DA) have been a subject of debate for several decades, due in part to the bewildering variety of processes in which it participates. DA plays diverse roles in reinforcement learning and action selection [1–9], motor control [10–12], vision [13, 14], interval timing [15–18], and attention [19]. Furthermore, DA functions through at least two channels: fast-timescale (phasic) signals and slow-timescale (tonic) signals [20]. While a large body of evidence has shown that phasic DA corresponds remarkably well with a ‘reward prediction error’ in reinforcement learning models [1–3, 21, 22], the functions of tonic DA, which span the domains above, remain unclear. Does this diversity of function reflect a heterogeneous computational repertoire? Or are there common principles that govern the role of tonic DA across these domains?

Across both experiments and theory, tonic DA is most widely studied in the domains of reinforcement learning and decision making. Experiments with Parkinson’s patients and healthy subjects have suggested that, when learning the associations between actions (or stimuli) and rewards, high DA facilitates learning from positive feedback, and low DA facilitates learning from negative feedback across a variety of tasks [4, 5, 23–27]. On the other hand, when this learned information must subsequently be used to select actions, DA seems to control the exploration-exploitation trade-off: Here, high DA promotes exploitation of actions with higher learned value and increases motivation. Low DA, on the other hand, promotes exploration [28–30] (but see [31, 32] and [Discussion](#)) and decreases motivation [33–35]. Recently developed computational models allow DA to achieve both learning and performance roles by endowing it with separate computational machinery for each [27, 36]. However, whether and how the effects of DA during learning and performance may be related at an algorithmic level remain open questions.

Another well-studied role of DA relates to its influence on interval timing. Much like in reinforcement learning, DA also seems to have two broad effects here. First, DA seems to modulate the speed of the internal timing mechanism: Acutely increasing DA levels results in behaviors consistent with a faster ‘internal clock,’ and acutely decreasing DA levels results in a slower internal clock [15–17, 37, 38]. In addition, unmedicated Parkinson’s patients, who are chronically DA-depleted, exhibit a distinct timing phenomenon known as the ‘central tendency’: When these patients learn intervals of different durations, they tend to overproduce the shorter intervals and underproduce the longer intervals [39–41]. While the central

tendency has been observed in healthy subjects [42–46] and animals [47], it is most pronounced in unmedicated Parkinson's patients, and DA repletion in these patients rescues veridical timing [39]. How the effects of DA on the speed of the internal clock and the central tendency are related, if at all, remains unclear.

Reinforcement learning and interval timing, though related at a theoretical and neural level [22, 48], have, until recently, largely been treated separately [49–51]. This has led to separate theories of DA developing for each: In reinforcement learning, an influential hypothesis has posited that tonic DA reflects average reward availability in a given context [34]. In other domains, however, tonic DA's role has been best explained as reflecting precision within a Bayesian framework [52], which we discuss below. The view that DA modulates precision (operationalized as the signal-to-noise ratio) has empirical grounding in interval timing [53–57], motor control [58], vision [13, 14], and audition [59]. Interestingly, beyond *true* precision, DA has also been associated with *estimated* precision, or the agent's estimate of its actual precision (or its confidence), independently of any changes in true precision [13, 60–62]. This is an important distinction, because a mismatch between true and estimated precision results in an underconfident or overconfident agent, which may affect behavior. In sum, the dual role of DA in true and estimated precision—and under what conditions this role holds—remains elusive. Once more, the duality of DA is not well understood.

Inspired by recent attempts to integrate the domains of reinforcement learning and interval timing [49–51], we will begin by introducing the concept of 'rational inattention,' borrowed from behavioral economics [63–65]. We will show that this framework unifies the two influential, yet seemingly distinct, algorithmic theories of tonic DA. We will then show that this framework predicts various empirical phenomena of reinforcement learning and interval timing, which cannot be explained by either theory alone.

Results

DA and rational inattention

To show how our account unites the two theories of tonic DA, let us begin by describing each one independently. First, under the average reward theory, tonic DA reports average reward availability in the current context. This theory has its roots in the observation that high tonic DA levels promote vigorous responses and high response rates in reinforcement learning tasks [34]. For a theoretical underpinning to this empirical phenomenon, Niv et al. [34] argued that animals in high-reward contexts should capitalize on the abundance of rewards with much of the same behaviors observed in hyperdopaminergic animals (high response vigor and high response rates). They thus proposed that tonic DA provides the link between average reward availability and the animal's behavioral response, i.e., tonic DA levels report average reward. More concretely, in this view, DA can be thought of as reporting the opportunity cost of 'sloth,' or the loss incurred by not capitalizing on the available rewards (in high-reward contexts, every passing moment not spent collecting reward is a wasted opportunity). Given that sloth is energetically inexpensive (and thus appealing), the average reward theory predicts that the animal will occupy this motivational state under low DA conditions (no use spending energy when reward potential is small), but will be increasingly incentivized to act quickly and vigorously as DA increases. The relationship of DA with responsivity and motivation has indeed been well documented [66–80].

Under the Bayesian theory, on the other hand, tonic DA signals the precision with which internal or external cues are stored and represented [52]. Thus under high DA, signaling high precision, the animal is more confident in its stored representations compared to the contextual information, and relies on them more heavily during decision making. This increased

reliance can be probed by examining conditions under which the cue representations are put in conflict with other sources of information. Most commonly, this entails comparing the ‘bottom-up’ information (e.g., sensory cues and their stored representations) with ‘top-down’ information (e.g., prior beliefs about the cues, based on the context): Under high DA, the animal weights bottom-up information more heavily, whereas low DA promotes top-down information. In Bayesian terms, which we describe explicitly in the next section, DA increases the animal’s estimate of the likelihood precision relative to the prior precision. This theory has been used to explain behavioral aspects of DA-related pathologies such as autism [81–84], schizophrenia [85–87], and Parkinson’s disease [41].

In thinking about the Bayesian theory of DA, it is important to distinguish between the *estimated* precision (what the agent perceives its precision to be) and *true* precision (what its precision actually is). True precision increases through an increase in signal-to-noise ratios, which improves performance. On the other hand, an increase in estimated precision, without an equal increase in true precision (unwarranted confidence), can actually *impair* performance. Recent work has shown that, depending on the circumstance, DA can influence true precision, estimated precision, or both, such as in interval timing [41, 53–57], motor control [58], vision [13], and memory [88]. However, why DA would freely modulate estimated precision independently of true precision, in the first place, is not well understood. After all, under normal circumstances, precision miscalibration is maladaptive (see ‘Precision miscalibration’ section).

Each theory outlined above succeeds in explaining a subset of empirical findings, as we will show. But how can tonic DA reflect both average reward availability and precision? We unite these two theories of DA under the framework of rational inattention, inspired by ideas originally developed in economics [63–65]. Rational inattention posits that cognitive resources are costly, and therefore will only be spent when the agent is incentivized to do so. In particular, ‘attention’ to a stimulus is the cognitive process through which the agent reduces its uncertainty about the stimulus (see [89] for related definitions), and can be formalized in terms of precision: With increased precision, the agent will have greater certainty about its environment, which will increase its ability to accumulate rewards, but it will also incur a greater cognitive cost. Consider, then, a task in which an agent can improve its performance by attending to certain stimuli. As average reward availability increases, the agent will be increasingly incentivized to pay the cognitive cost necessary to increase precision. In this way, rational inattention provides the critical link between average reward and precision, and we hypothesize that this coupling is instantiated through DA: By reporting average reward, DA determines precision.

In the next sections, we will formalize this simple intuition and show how it allows us to expand the scope of experimental predictions made by each theory individually, while also conceptually connecting a number of seemingly distinct views of tonic DA.

Bayesian inference

Before presenting the model, let us briefly discuss Bayesian inference, which will help us formalize the notion of precision (both true and estimated) and its effects on performance.

Suppose an animal is learning some parameter μ . For instance, μ may represent the expected reward obtained from some reward source, or the temporal interval between an action and an outcome. Because various stages of this learning process are characterized by noise [90], beginning with the nature of the parameter itself (stimuli are seldom deterministic) and leading up to the storage process (neurons are noisy), the learned information can be represented by a distribution over parameter values. This is referred to as a likelihood function

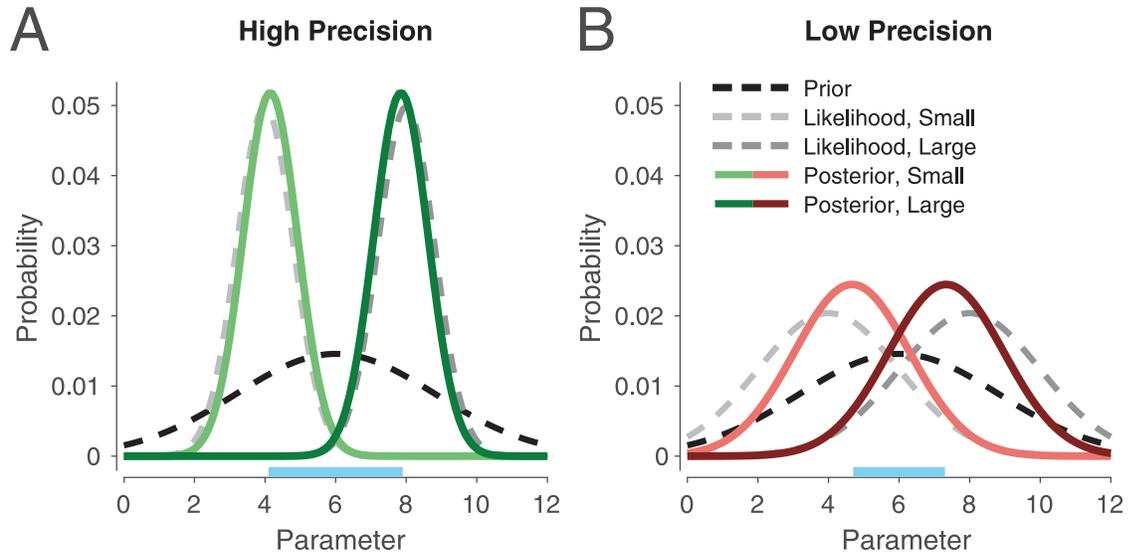


Fig 1. Illustration of Bayesian inference for a task with two parameters of different magnitudes, one small and one large, under either high or low precision. The posterior is the normalized product of the prior and likelihood. (A) When likelihood precision is high compared to prior precision, the posterior will remain close to the likelihood. (B) As the ratio of likelihood precision to prior precision decreases, the posterior migrates toward the prior. Note here that likelihood precision controls the distance between the posterior means (compare lengths of blue segments on the x-axis), a point we return to later.

<https://doi.org/10.1371/journal.pcbi.1008659.g001>

(gray curves in Fig 1; light and dark gray curves represent likelihood functions for a parameter with small and large magnitude, respectively). The spread of this distribution reflects the precision, or reliability, of the encoding process: High precision will lead to a tight distribution (Fig 1A), whereas low precision will lead to a wide distribution (Fig 1B). For simplicity, we take these distributions to be Gaussian throughout.

During subsequent decision making, the animal must use this likelihood function to produce a learned estimate of μ , which we denote by $\hat{\mu}$. An optimal agent will use all the information available to it in order to produce this estimate, including its knowledge of the current context, or the ‘prior distribution’ (black curves in Fig 1). For instance, if reward sources in this context tend to yield relatively larger rewards, or intervals tend to be longer, then it makes sense for the animal to produce an estimate that is slightly larger than the likelihood mean, especially when the precision of the likelihood is low (i.e., when it is not very reliable; compare black curves with light gray curves). Bayes’ rule formalizes this intuition [91–95] and states that an optimal agent will take the product of the likelihood and prior to compute the posterior distribution over parameter values (green and red curves in Fig 1):

$$p(\mu|m) \propto p(m|\mu) p(\mu), \tag{1}$$

where m represents the stored (subjective) values, $p(m|\mu)$ is the likelihood, $p(\mu)$ is the prior, and $p(\mu|m)$ is the posterior. Under standard assumptions for Gaussian distributions [43, 95–97], the estimate $\hat{\mu}$ obtained by a Bayesian ideal observer will correspond to the posterior mean, which can be computed using the two quantities characterizing each distribution, their means and precisions:

$$\hat{\mu} = \left(\frac{\lambda_0}{\lambda_0 + \lambda} \right) \mu_0 + \left(\frac{\lambda}{\lambda_0 + \lambda} \right) \mu. \tag{2}$$

Here, μ_0 , λ_0 , μ , and λ represent the prior mean, prior precision, likelihood mean, and

likelihood precision, respectively, in expectation. In words, Eq 2 states that the posterior mean $\hat{\mu}$ is a weighted average of the prior mean μ_0 and the likelihood mean μ , and their respective precisions λ_0 and λ constitute the weights after normalization. Hence, the tighter each distribution, the more it pulls the posterior mean in its direction (compare Fig 1A and 1B). This is consistent with the intuition that one should rely more on what is more reliable. This type of optimal computation has been observed in many different domains [98].

In summary, when incoming information is noisy (low precision), an optimal agent will more strongly modulate its responses based on context. On the other hand, when the agent is very confident in incoming information (high precision), it will show little influence by context.

Model description

Consider now a task where an animal must learn certain parameters μ in order to maximize its expected reward. For instance, in a temporal task, a rat may be trained to wait for a fixed period before pressing a lever in order to receive a drop of sucrose. Alternatively, in a reinforcement learning task, it may need to learn the magnitudes of rewards delivered from two different sources in order to choose the source with higher expected reward in the future. We can model these problems with the general encoding-decoding framework shown at the bottom of Fig 2. Here, the animal transforms the objective stimulus μ —the duration of a timed interval or the magnitude of a reward—into a likelihood distribution $p(m|\mu)$ (encoding), which, during the performance stage, it then uses to produce its estimate $\hat{\mu}$ of the original stimulus via Bayes’ rule (decoding).

This framework sheds light on the distinction between true and estimated precision. In the previous section, we implicitly assumed that the *encoding* precision λ is faithfully relayed to the *decoding* stage for Bayesian inference (i.e., we assumed the same λ for both stages). However,

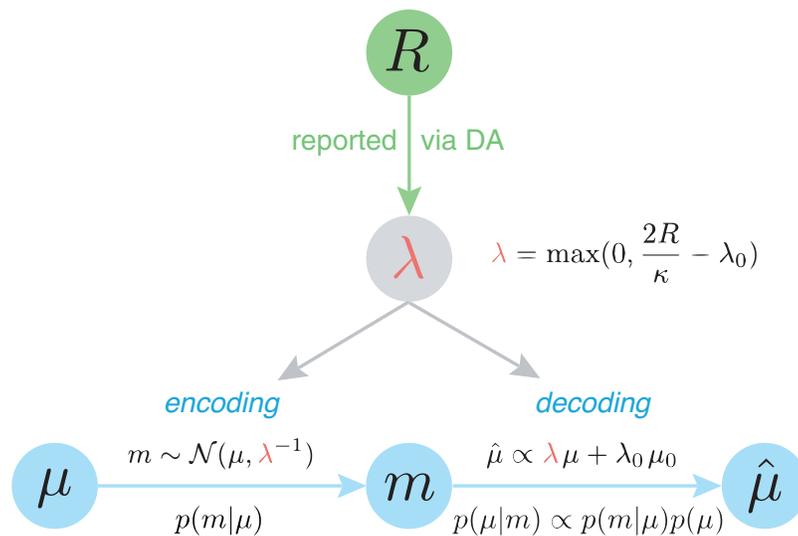


Fig 2. Rational inattention model of DA. Under rational inattention, average reward controls the likelihood precision through tonic DA. As derived in the next section and summarized in the top equation, increases in average reward increase likelihood precision, which in turn affects both encoding and decoding. Because average reward is a property of the context, DA can relay the likelihood precision (i.e., the precision of encoding) to the decoding stage, even when encoding and decoding are temporally decoupled. R : reward; κ : unit cost of attention (or information; see next section); λ : likelihood precision; λ_0 : prior precision; μ : likelihood mean (objective values); m : subjective (stored) values; μ_0 : prior mean; $\hat{\mu}$: posterior mean (estimate of μ); $p(m|\mu)$: likelihood; $p(\mu)$: prior; $p(\mu|m)$: posterior.

<https://doi.org/10.1371/journal.pcbi.1008659.g002>

increasing λ selectively during decoding will increase the agent's estimated, but not its true, precision. We refer to this as 'precision miscalibration.' Here, Bayesian inference will overestimate the precision with which encoding occurred, resulting in an excessive reliance on the likelihood compared to the prior (Fig 1). Such miscalibration is maladaptive, and in a subsequent section we show that optimal performance occurs when the same precision used during encoding is also used during decoding.

Empirical evidence supporting the Bayesian theory suggests that DA influences both the true and estimated precision. But then, how is precision faithfully maintained across stages? After all, encoding and decoding can be temporally separated, and, given the natural fluctuations of tonic DA, what is the guarantee that the DA level would be the same at both stages?

We hypothesize that tonic DA reports the single quantity of average reward in a given context. According to rational inattention, this determines the precision λ . Imagine then that an animal experiences a high-reward context, learns its average reward, then leaves that context. If at a later date, the animal is returned to that same context, then its behavior there will depend on the average reward that it had previously learned, which is encoded in the DA signal. This means that the appropriate estimated precision can be retrieved from the DA signal when in that context. In this manner, by reporting average reward, DA during the encoding stage controls the true precision, and during the decoding stage, determines a faithful estimate of this true precision.

Of note, this soft 'guarantee' against precision miscalibration breaks down when DA levels are manipulated directly: For instance, while generally high levels of DA will increase both the agent's true and its estimated precision, increasing DA selectively during decoding will increase the agent's estimated, but not its true, precision. In the next sections, we will relate this miscalibration to experimental studies in which DA was acutely manipulated during the decoding stage. We will also show how this framework subsumes the view that DA implements 'gain control' on action values during action selection, without having to posit additional computational machinery.

In summary, by communicating the single quantity of context-specific average reward during both encoding and decoding, DA can achieve two seemingly distinct functions—controlling the precision of encoding incoming information as well as modulating the reliance on previously learned information when executing decisions.

Mathematical description of rational inattention

Rational inattention views the agent as weighing the rewards gained from increased precision against the cost of this increase. With too much of an increase in precision, the cognitive cost may outweigh the additional rewards, whereas with too little of an increase in precision, the agent may forgo uncanceled potential rewards. Let us begin by formalizing these two terms. This will then allow us to characterize the precision needed to maximize the agent's utility (rewards minus cost).

In this section, we assume a perfectly calibrated agent, and hence make no distinction between estimated and true precision. The effects of precision miscalibration will be examined in the next section.

Consider an objective stimulus μ , such as a sensory cue or an environmental variable, that the agent must estimate. Let m denote the signal actually encoded by the agent, which is subject to noise that is reducible at a cost. Neurally, this may refer to noise at any level from signal transduction to storage, causing the stored neural representation m to be different from μ . We can model m as being drawn from a Gaussian distribution with mean μ and precision λ (the likelihood distribution), i.e., $m \sim \mathcal{N}(\mu, \lambda^{-1})$, where the agent has control over λ (Fig 2). The

latent variable μ (the ‘source’) is drawn from some prior distribution representing the context, which we also take to be Gaussian for simplicity, with mean μ_0 and precision λ_0 , i.e., $\mu \sim \mathcal{N}(\mu_0, \lambda_0^{-1})$. Then by Eq 2, the posterior mean can be written as

$$\hat{\mu}(m) = wm + (1 - w)\mu_0, \tag{3}$$

where

$$w = \frac{\lambda}{\lambda + \lambda_0} \tag{4}$$

is the weight given to the likelihood mean, bounded by 0 and 1. We can now compute the average error in estimating the source. Intuitively, when λ is small, the estimate will migrate toward the prior mean, causing a large mismatch between the source and its estimate, whereas when λ is large, this migration—and hence the error—is small. For analytical convenience, we quantify this error using a quadratic error function:

$$E(\mu, \hat{\mu}(m)) = (\mu - \hat{\mu}(m))^2. \tag{5}$$

Using standard Gaussian identities, the ‘conditional’ error is given by

$$\begin{aligned} \mathbb{E}[E(\mu, \hat{\mu}(m))|\mu] &= \int_m E(\mu, \hat{\mu}(m))\mathcal{N}(m; \mu, \lambda^{-1})dm \\ &= w^2\lambda^{-1} + (1 - w)^2(\mu - \mu_0)^2. \end{aligned} \tag{6}$$

The conditional error can be understood as the agent’s expected response variance for a particular source. To compute the overall response variance (the ‘marginal’ error), we average across all sources:

$$\begin{aligned} \mathbb{E}[E(\mu, \hat{\mu}(m))] &= \int_{\mu} \mathbb{E}[E(\mu, \hat{\mu}(m))|\mu]\mathcal{N}(\mu; \mu_0, \lambda_0^{-1})d\mu \\ &= w^2\lambda^{-1} + (1 - w)^2\lambda_0^{-1} \\ &= (\lambda + \lambda_0)^{-1}. \end{aligned} \tag{7}$$

Thus an increase in the encoding precision λ decreases the marginal error, which in turn improves performance.

Next, let us formalize the cognitive cost of collecting information (reducing uncertainty) about the environment to improve performance. An analytically tractable choice of attentional cost function is the mutual information [99, 100], which measures the expected reduction in uncertainty due to the observation of the signal m :

$$\begin{aligned} I(\mu; m) &= H(\mu) - \mathbb{E}[H(\mu|m)] \\ &= \frac{1}{2} \log \left(\frac{2\pi e}{\lambda_0} \right) - \frac{1}{2} \log \left(\frac{2\pi e}{\lambda_0 + \lambda} \right) \\ &= \frac{1}{2} \log \left(1 + \frac{\lambda}{\lambda_0} \right), \end{aligned} \tag{8}$$

where $H(\mu)$ denotes the entropy of the probability distribution of μ . Intuitively, the uncertainty about μ , before observing the signal, can be quantified as the entropy of the prior distribution (high entropy, or uncertainty). On the other hand, after observing the signal, the uncertainty is represented by the entropy of the posterior distribution (lower entropy). The mutual information measures this reduction, which rational inattention assumes comes at a cost. The second

equality follows from the entropy of Gaussian distributions with precision λ_0 (the prior distribution; first term) and precision $\lambda_0 + \lambda$ (the posterior distribution; second term). Note that the posterior precision is simply the sum of the prior and likelihood precisions.

The choice of mutual information is appealing because it is tractable, displays the correct behaviors (increases in mutual information increase precision) [101], and is not restricted to a particular neural implementation. This final point is important given that the biological basis of cognitive effort remains unclear, despite attempts to link the two [102]. Indeed, while we assume the agent is primarily concerned with the cost of collecting and storing units of information, the rational inattention framework remains agnostic to the source of this cost at a biological level.

Assuming an increase in precision can improve performance, which for convenience we take to be linear with accumulated rewards, we can now combine the attentional cost function with the error function to define the attentional optimization problem:

$$\min_{\lambda} R \mathbb{E}[E(\mu, \hat{\mu}(m))] + \kappa I(\mu; m), \quad (9)$$

where R is the reward incentive, which we propose is reported by DA, and $\kappa > 0$ is the unit cost of information. As an incentive, R formally refers to the subjective, rather than the objective, value of the reward. This distinction will not be important for our model unless the two are specifically dissociated (e.g., through changes to satiety or baseline reward availability; see ‘DA and controllability’ section).

The agent seeks to minimize both the performance error and the (costly) reduction of uncertainty, which are weighted by the reward incentive and the unit cost of information, respectively. The idea here is that a unit increase in information decreases error (which leads to higher utility) but increases costs (which leads to lower utility). Thus if the agent pays *too* much attention to the task, the costs may outweigh the benefits, whereas if the agent pays *no* attention to the task, it may not reap any rewards. For our choice of cost and error functions, a middle ground exists that optimizes the agent’s utility (i.e., solves the optimization problem):

$$\lambda^* = \max\left(0, \frac{2R}{\kappa} - \lambda_0\right). \quad (10)$$

The rational inattention solution has three intuitive properties: First, attention to the signal increases with reward incentive (R). Second, attention to the signal decreases with the cost of information (κ). Third, attention to the signal decreases with prior precision (λ_0). In other words, if the agent is more confident about the source before observing the signal, it will pay less attention to the signal. After all, there is no need to spend energy gathering information about a source when that source is already well known (Fig 3).

It is important to note that one can construct cost and error functions where such a middle ground is not attainable. For example, if the information cost and performance benefit take exactly equal but opposite shape, then the agent should always either increase its precision to infinity (if $R > \kappa$) or decrease it to zero (if $R < \kappa$). Our choice of functions, while principled, primarily serve to capture the intuition of the rational inattention framework.

Having derived the relationship between reward incentive and precision in Eq 10, let us briefly examine how it is affected by the prior precision and information cost. The relationship in Eq 10 is piecewise linear (Fig 3): When the incentive R is small, the precision λ^* is zero, and only after R becomes sufficiently large does λ^* begin to increase. Intuitively, the agent’s performance depends on its posterior precision, which is the sum of its prior and likelihood precisions. Under rational inattention, the agent seeks to ‘match’ its posterior precision with the reward incentives (divided by the information cost). When the prior precision alone

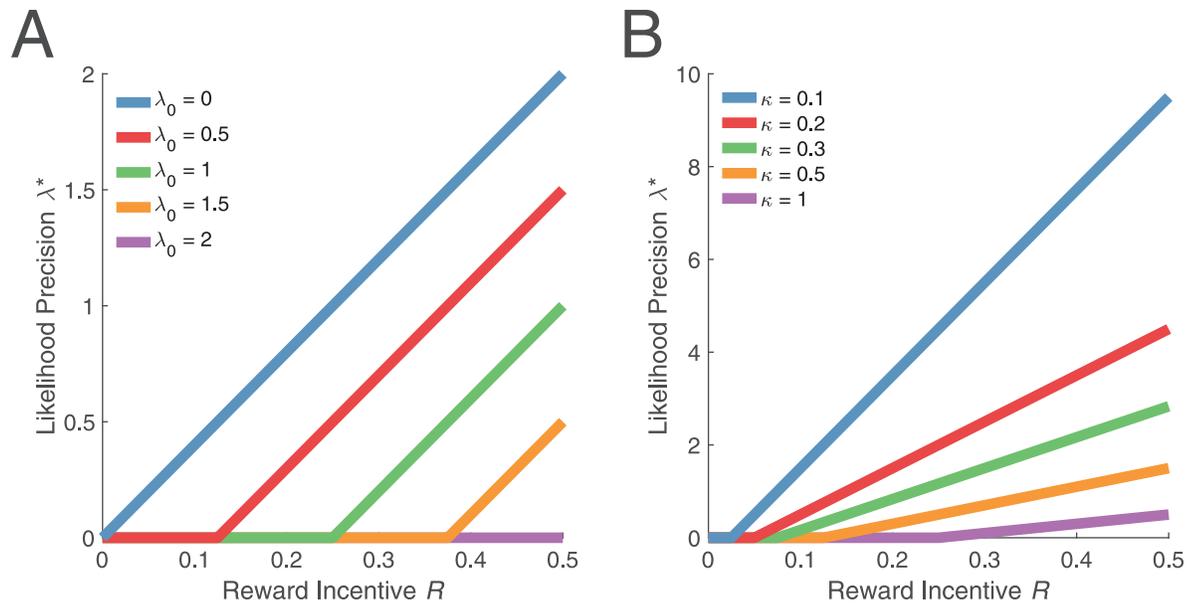


Fig 3. Relationship between reward incentive and likelihood precision under different levels of prior precision and information cost. The relationship between λ^* and R is piecewise linear: When R is small, the agent is not sufficiently incentivized to attend to the task and relies only on its prior knowledge. When R is sufficiently large, the agent linearly increases λ^* with R . (A) Increases in λ_0 shift the piecewise linear function to the right. (B) Increases in κ shift the function to the right and decrease the slope. For illustration, we have fixed $\kappa = 0.5$ and $\lambda_0 = 0.5$ in (A) and (B), respectively.

<https://doi.org/10.1371/journal.pcbi.1008659.g003>

achieves—or exceeds—the optimal posterior precision given the reward incentive, there is no need to attend to the task and collect new (costly) information, so λ^* is simply zero (horizontal segment of piecewise linear function). On the other hand, when the reward incentive calls for a higher posterior precision than the prior precision, the agent should make up the difference by attending to the task (increasing ray of piecewise linear function). The point at which the two pieces of the function intersect corresponds to the value of λ_0 that exactly matches the reward incentive.

It is straightforward to consider how changes in λ_0 and κ affect the relationship between R and λ^* . For larger values of λ_0 , the agent must be incentivized more before it begins to attend to the task (the point at which λ^* begins to increase shifts to the right). But after this point, a unit increase in R will increase λ^* by the same amount regardless of λ_0 (same slope of increasing ray with different λ_0 ; Fig 3A). Similarly, for larger values of κ , larger reward incentives R will be needed for the agent to attend to the task. However, because precision depends on the ratio between R and κ , a unit increase in R will have a weaker effect on λ^* when κ is large (shallower slope; Fig 3B).

Finally, it should be noted that rate-distortion theory offers a normative interpretation of our optimization problem: If we accept that there is an upper bound on the bit rate of perception, then optimizing reward subject to this bit rate constraint will lead to Eq 9 (the Lagrangian function), a standard result from rate-distortion theory [103].

Precision miscalibration

Let us now analyze the effects of miscalibrated precision on the accuracy of a Bayesian agent. With λ and λ' denoting true and estimated precision, respectively, we set $c = \lambda'/\lambda$, which can be thought of as a miscalibration factor: When $c < 1$, precision is underestimated, and when $c > 1$, precision is overestimated. Here we examine the error incurred by miscalibration.

Following the derivations in the previous section, the signal weight can now be written as

$$w = \frac{c\lambda}{c\lambda + \lambda_0}, \quad (11)$$

and the marginal error is

$$\mathbb{E}[E(\mu, \hat{\mu}(m))] = w^2\lambda^{-1} + (1-w)^2\lambda_0^{-1}. \quad (12)$$

Taking the partial derivative of this expression with respect to c and setting it to 0, we find, consistent with intuition, that the error-minimizing miscalibration factor is $c = 1$. Thus, as an example, any experimental manipulation that increases estimated precision λ' without increasing true precision λ will produce a miscalibration factor greater than 1 and thereby incur an increase in error. Intuitively, a miscalibration factor of $c > 1$ means that both the signal *and* noise are amplified.

Note here that ‘miscalibration’ only refers to differences between true and estimated precision, and not to differences between the stimulus μ (or likelihood) and estimate $\hat{\mu}$ (or posterior). It is possible for μ and $\hat{\mu}$ to be different without precision miscalibration, as shown in the previous section.

Relationship with experimental data

DA and reinforcement learning. Let us now examine the predictions of our framework for the special case of reinforcement learning. Here, the task is to learn reward magnitudes and subsequently select the appropriate actions.

Our first prediction is that high DA during decoding will silence contextual influence, or, in Bayesian terms, decrease the influence of the prior. As illustrated in [Fig 1](#), an increase in likelihood precision, via high DA, amplifies the difference between the posterior means (compare blue horizontal segments on the x-axis). The parameter being reward magnitude here, this amplification makes the agent more likely to choose the action with the higher reward, i.e., to exploit. Similarly, low DA leads to low estimated precision and a significant Bayesian attraction toward the prior, which decreases the difference between the posterior means and promotes exploration ([Fig 4B](#)). Indeed, previous work has suggested that DA controls the exploration-exploitation trade-off, whereby high DA encourages exploiting the option with the highest reward, and low DA encourages exploring other options [[28–30](#)] (but see [[31](#)] and [Discussion](#)). For instance, Cinotti et al. [[30](#)] trained rats on a non-stationary multi-armed bandit task with varying levels of DA blockade. The authors observed that the degree of win-shift behavior, representing the drive to explore rather than to exploit, increased with higher doses of the DA antagonist flupenthixol ([Fig 4A](#)). Previous reinforcement learning theories of DA have explained this finding by suggesting that, during action selection, DA mediates gain control on action values [[27](#), [104](#)]. In the [Methods](#), we show that DA’s decoding effect in our framework is equivalent to controlling the temperature parameter in the softmax function, a standard form of gain control [[105](#), [106](#)]. Thus, rational inattention endogenizes DA’s role in exploitation, without needing to posit a separate gain control mechanism for DA that only appears during performance.

Our second prediction concerns the modulation of learning by DA. Learning can be thought of as iteratively weighing incoming information against a previously learned estimate to produce a new estimate. In so doing, the animal also learns the distribution of stimuli, which allows it to construct a prior for its context, as in [Fig 1](#). Under our framework, high DA signals high average reward in the context. Therefore, an agent under high DA should expect—and thus initialize its prior at—high rewards. This will result in faster learning of high

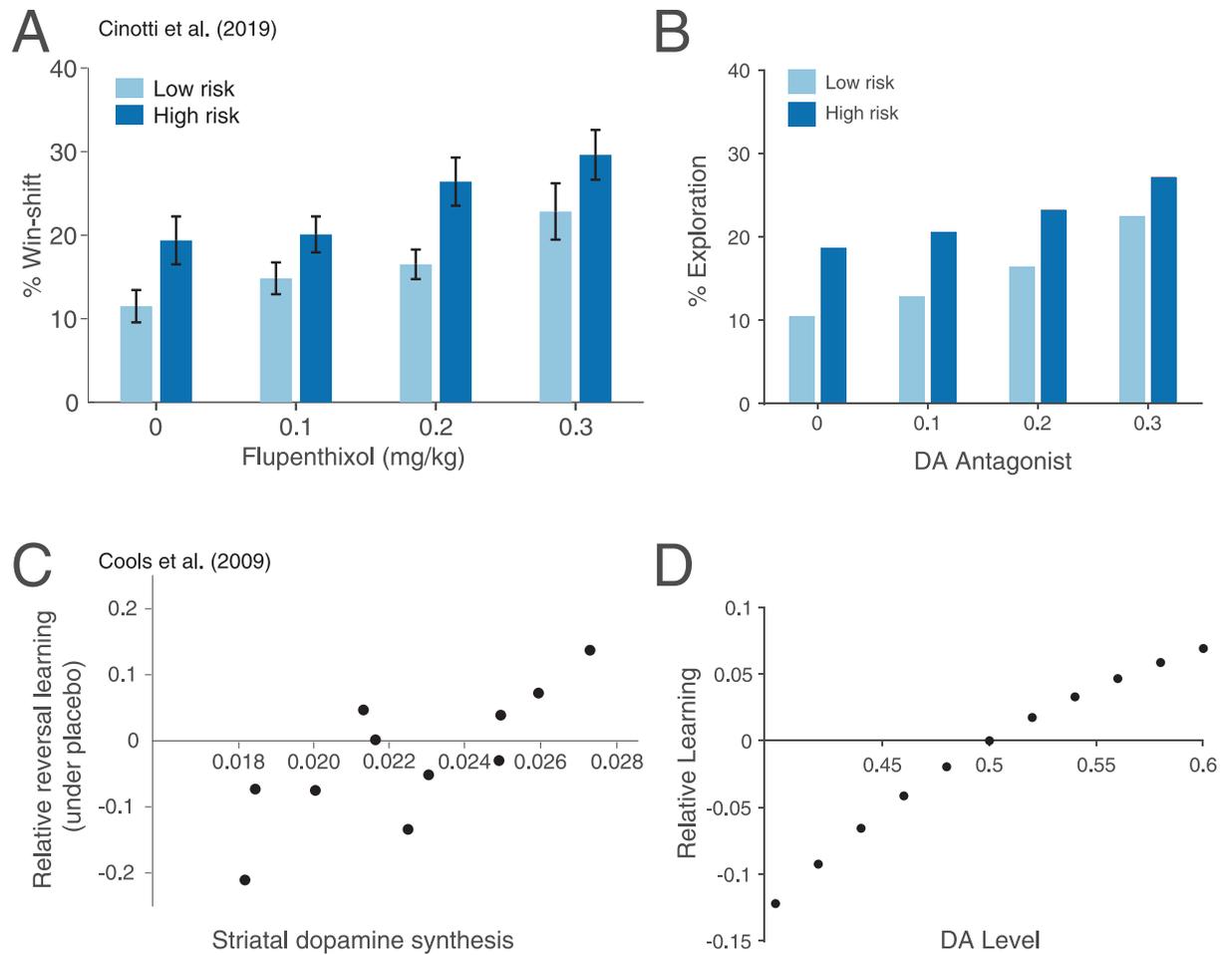


Fig 4. Rational inattention and reinforcement learning. (A) Using a non-stationary three-armed bandit task, Cinotti et al. [30] have shown that the DA antagonist flupenthixol promotes exploration (win-shift behavior). The task furthermore included two different risk levels: In low-risk conditions, one lever was rewarded with probability 7/8, and the other two with probability 1/16 each; in high-risk conditions, one lever was rewarded with probability 5/8, and the other two with probability 3/16 each. The effect of DA was evident for both conditions, and win-shift behavior was more pronounced in high-risk conditions across DA levels. Figure adapted from [30]. (B) Our model recapitulates these results: As DA decreases, likelihood precision decreases, which in turn reduces the difference in posterior means, and encourages exploration. The effect of risk on exploration follows from the reduced differences in likelihood means (and, by extension, posterior means) in high-risk compared to low-risk conditions. (C) Cools et al. [23] have shown that human subjects with high DA synthesis capacity learn better from unexpected rewards than from unexpected omissions of reward, whereas subjects with low DA synthesis capacity learn better from unexpected omissions than unexpected rewards. Relative accuracy is the accuracy following unexpected rewards minus the accuracy following unexpected punishments, taken to indicate the extent of learning from positive feedback compared to negative feedback. Figure adapted from [23]. (D) Our model recapitulates this result: High DA shifts the agent's beliefs toward expecting positive feedback. Thus the speed of convergence for learning high rewards increases. Similarly, the speed of convergence for learning low rewards increases under low DA. The asymmetry in relative learning is due to an asymmetry in precision: Under high DA, likelihood precision is high, so the prior has a weaker effect than under low DA. For (B, D), see [Methods](#) for simulation details.

<https://doi.org/10.1371/journal.pcbi.1008659.g004>

rewards compared to low rewards, or equivalently, of positive feedback (rewards higher than expected) compared to negative feedback (rewards lower than expected). Similarly, under low DA, negative feedback will be learned better than positive feedback (Fig 4D). Indeed, tonic DA levels have been shown to control the relative contribution of positive and negative feedback to learning [4, 5, 23, 107]. For instance, Cools et al. [23] repeatedly presented human subjects with a pair of stimuli (images), where one stimulus was associated with reward, and the other with punishment. On each trial, one of the stimuli was highlighted, and subjects had to predict

whether that stimulus would lead to reward or punishment. Occasionally, unsignaled reversals of the stimulus-outcome contingencies would occur, so that the first stimulus to be highlighted after the reversal would result in an unexpected outcome. The same stimulus would then be highlighted on the subsequent trial. Accuracy on this second trial reflected the extent to which subjects learned from unexpected rewards (if the previously punished stimulus was highlighted) vs. unexpected punishments (if the previously rewarded stimulus was highlighted). The authors showed that subjects with higher DA synthesis capacity learned better from unexpected rewards, whereas those with lower DA synthesis capacity learned better from unexpected punishments (Fig 4C). Interestingly, under rational inattention, learning better from positive or negative feedback is not a bias, but rather an optimal strategy.

It should be noted here that while striatal DA synthesis may in principle affect both phasic and tonic levels, the results of Cools et al. [23] cannot be explained as simply amplifying phasic DA, which putatively encodes reward prediction errors, without affecting tonic DA. For instance, recent work has shown that synthesis capacity and striatal prediction errors are in fact negatively correlated [108]. Furthermore, differential learning by positive vs. negative feedback has also been observed using pharmacological manipulations, which affect tonic DA directly [4].

DA and interval timing. We now examine our framework for the case of interval timing. We will focus on reproduction tasks, in which subjects must produce a previously learned interval under different manipulations, although our predictions will apply equally well to discrimination tasks, in which subjects respond differently to intervals of different lengths (e.g., responding ‘short’ or ‘long’ depending on whether a new interval is shorter or longer than a previously learned one). For each reproduction result below, we model its discrimination counterpart in [S3 Appendix](#).

Our first prediction is that while timing under high DA levels will be nearly veridical, timing under low DA levels will be susceptible to interfering temporal stimuli (strong migration toward the prior; Fig 5B). Indeed, unmedicated Parkinson’s patients strongly display this effect, referred to as the central tendency. Here, reproducing durations of different lengths results in shorter intervals being overproduced and longer intervals being underproduced [39, 40], and veridical timing is rescued with DA repletion [39] (Fig 5A; see also [S3 Appendix](#)). Shi et al. [41] have shown that these behaviors conform remarkably well to a Bayesian framework in which DA modulates the precision of the likelihood. Rational inattention takes this one step further: Because DA reflects average reward, our framework also predicts the central tendency under low average reward conditions or satiated states (in which rewards are devalued). This is consistent with empirical studies that manipulated average reward and satiety through pre-feeding [109–111], although motivation is a confound in these experiments ([S4 Appendix](#)).

Our second prediction concerns acute manipulations of DA during decoding. A large body of work has shown that tonic DA affects interval timing by modulating the speed of the internal clock (or the ‘subjective time’), whereby higher DA levels lead to a faster clock [15–17, 37, 38, 112–116] (Fig 5C). This finding has been replicated under different experimental paradigms to control for potential confounds (e.g., motivation; [S3 Appendix](#)). Effects on clock speed are also well documented in the behavioral literature: Here, the speed of the clock increases when animals are placed in high reward-rate contexts and decreases in low reward-rate contexts [117–123] ([S4 Appendix](#)). Clock speed similarly decreases in satiated states (i.e., in reward devaluation) [124, 125].

Our framework predicts these findings under the assumption that temporal precision is controlled by the internal clock speed. Recent empirical findings provide some support for this assumption. Indeed, a number of studies have identified ‘time cells’ (e.g., in striatum and medial frontal cortex [126, 127]), which seem to function as an internal timing mechanism:

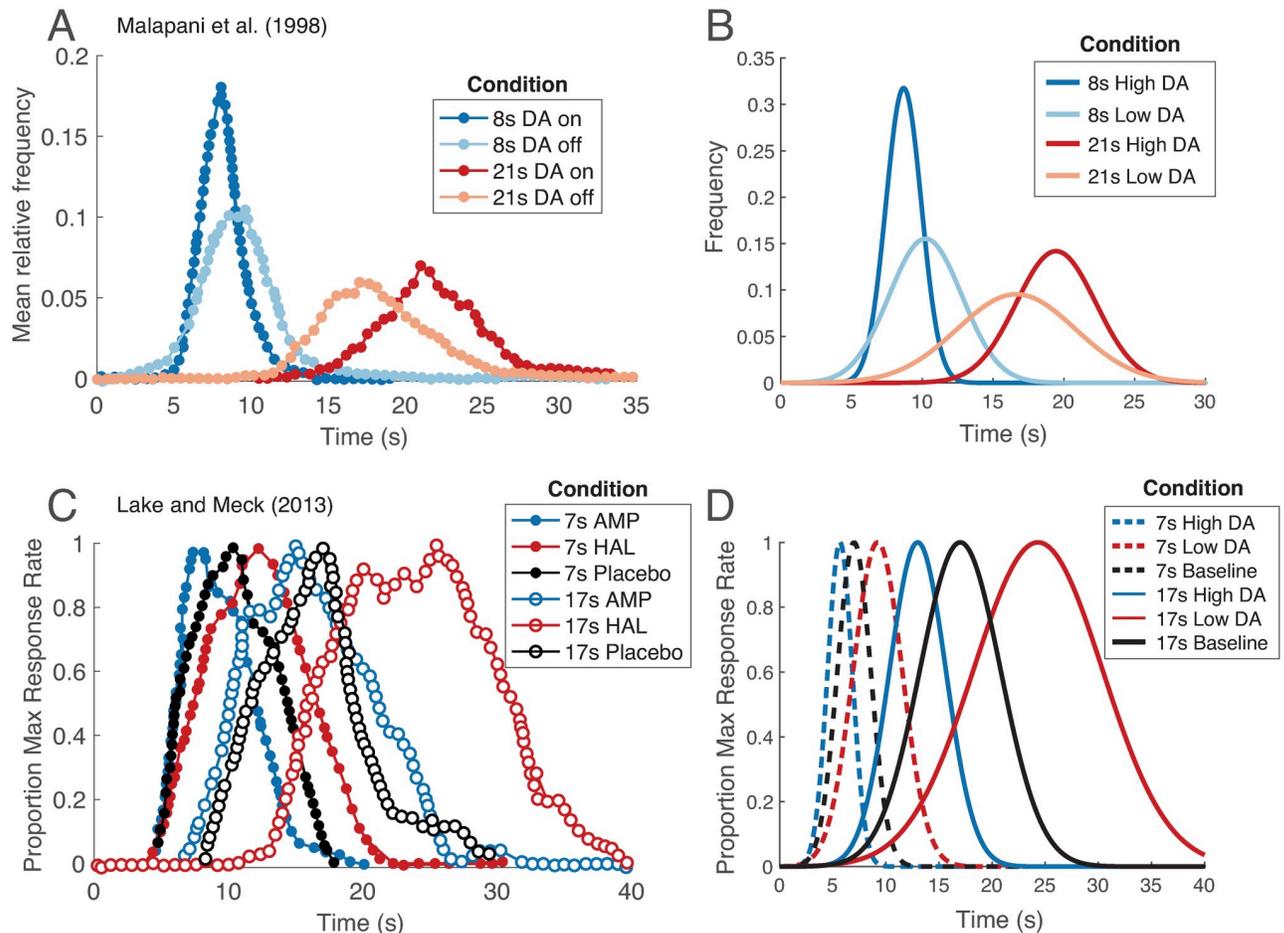


Fig 5. Rational inattention and interval timing. (A) DA masks the central tendency effect. Malapani et al. [39] have shown that when unmedicated Parkinson's patients learn intervals of different durations in an interleaved manner, they overproduce the shorter intervals and underproduce the longer ones (light blue and orange curves). Medication rescues veridical timing (dark blue and red curves). Figure adapted from [39]. (B) Our model recapitulates this effect: When DA is high, likelihood precision is high, and the posterior closely resembles the likelihood. When DA is low, likelihood precision is low, and the posterior migrates toward the prior. (C) DA increases the speed of the internal clock. Lake and Meck [17] trained healthy human subjects on reproducing a 7- or 17-second interval. They then acutely administered either amphetamine (DA agonist) or haloperidol (DA antagonist), and observed temporal reproduction that was consistent with either a faster or a slower clock, respectively. Note that the central tendency is not captured in this experiment because each of the 7-second and 17-second intervals was presented separately in a blocked manner, but plotted in the same figure for convenience. Figure adapted from [17]. (D) Our model recapitulates this effect: When DA is high, temporal receptive fields must compress against objective time to increase likelihood precision. This results in a faster internal clock. When DA is low, temporal receptive fields expand to decrease likelihood precision. This results in a slower internal clock. For (B, D), see [Methods](#) for simulation details.

<https://doi.org/10.1371/journal.pcbi.1008659.g005>

Time cells fire sequentially over the course of a timed interval, they tile the interval, and their activations correlate with timing behavior. In other words, their activations seem to reflect 'temporal receptive fields.' By definition, temporal precision is inversely related to the temporal receptive field width. Thus, any rescaling of the time cells will modify the precision but also change the speed of the internal clock (or change the mapping between objective and subjective time; [S2 Appendix](#)). Rescaling of time cell activations has indeed been well documented [126, 127].

Given that the clock speed can change, how is it then that timing can ever be reliable? Under rational inattention, the answer is straightforward: Reporting context-specific average reward, DA maintains the same precision across both encoding and decoding. If precision is

implemented through changes in the speed of the internal clock, it follows that the clock speed will also be the same during encoding and decoding, and, in general, timing will be reliable.

Let us now turn to acute manipulations of DA. Under rational inattention, an acute increase in DA levels at decoding increases estimated precision, implemented here by compressing the time cell activations against objective time (S2 Appendix). This increases the speed of the internal clock, which results in underproduction of previously learned intervals. Similarly, acutely decreasing DA levels at decoding will slow down the internal clock, resulting in overproduction of stored intervals, consistent with the empirical findings (Fig 5D). This framework also predicts the influence of average reward on clock speed (S4 Appendix) as well as that of different motivational states—Under rational inattention, reward and DA manipulations are equivalent.

Finally, note that the ability to keep track of time worsens as the interval duration increases. This worsening is known as Weber's law, which asserts that the inverse square root of precision—or the standard deviation—increases linearly with time [128–130]. The predictions specific to our model do not depend on Weber's law, whose underlying cause is still a subject of active debate [129, 131, 132] (only the pattern of wider probability distributions for larger intervals depends on Weber's law). However, for a completely determined mathematical model, we propose a rational-inattention-based derivation of this phenomenon in S1 Appendix. This will infuse our model with quantitatively precise predictions without affecting our qualitative results.

DA and controllability. Finally, rational inattention makes the counterintuitive prediction that the effects of average reward in increasing the speed of the clock will be conditional on controllability. In other words, attentional cost is only paid if outcomes can be improved. Otherwise, there is no use in spending cognitive resources on a task beyond the animal's control.

Bizo and White [133, 134] sought to directly examine the effect of freely delivered reinforcers on interval timing. To do so, they trained pigeons on a free-operant task that allowed for left-key responding and right-key responding. In each 50-second trial, only the left-key response was rewarded during the first 25 seconds, and only the right-key response was rewarded during the last 25 seconds. As expected, trained pigeons were more likely to select the left-key response early on, and the right-key response later on. This response was quantified with a psychometric function, in which the probability of making a right-key response was plotted against time, and which took a sigmoidal shape. To examine the effect of free rewards, the authors additionally included a center key that would freely deliver rewards independent of the timing task. From the psychometric function, the authors fit a computational model of timekeeping to deduce a 'pacemaker period' (inverse of clock speed). Intuitively, a shift in the psychometric function to the left (being more likely to select the right-key response early) was taken to indicate that the clock speed increased (smaller pacemaker period), whereas a shift to the right was taken to indicate that the clock speed decreased (larger pacemaker period). The authors found that the speed of the internal clock varied *inversely* with freely delivered rewards (Fig 6A), consistent with the principle of rational inattention: As freely delivered rewards became more available, the incentive to pay attention to the timing task decreased, causing precision, and thus clock speed, to decrease (Fig 6B; Methods). On the other hand, when performance-contingent rewards became more available compared to freely delivered rewards, the clock speed increased (Fig 6A and 6C). Experiments in mice suggest that DA levels are higher in tasks with controllable outcomes and lower under learned helplessness (lack of control) [135, 136], as predicted by our framework.

Should the clock speed decrease to zero in the complete absence of controllability? It is reasonable to assume here that, even in this case, the animal should still pay *some* attention to the

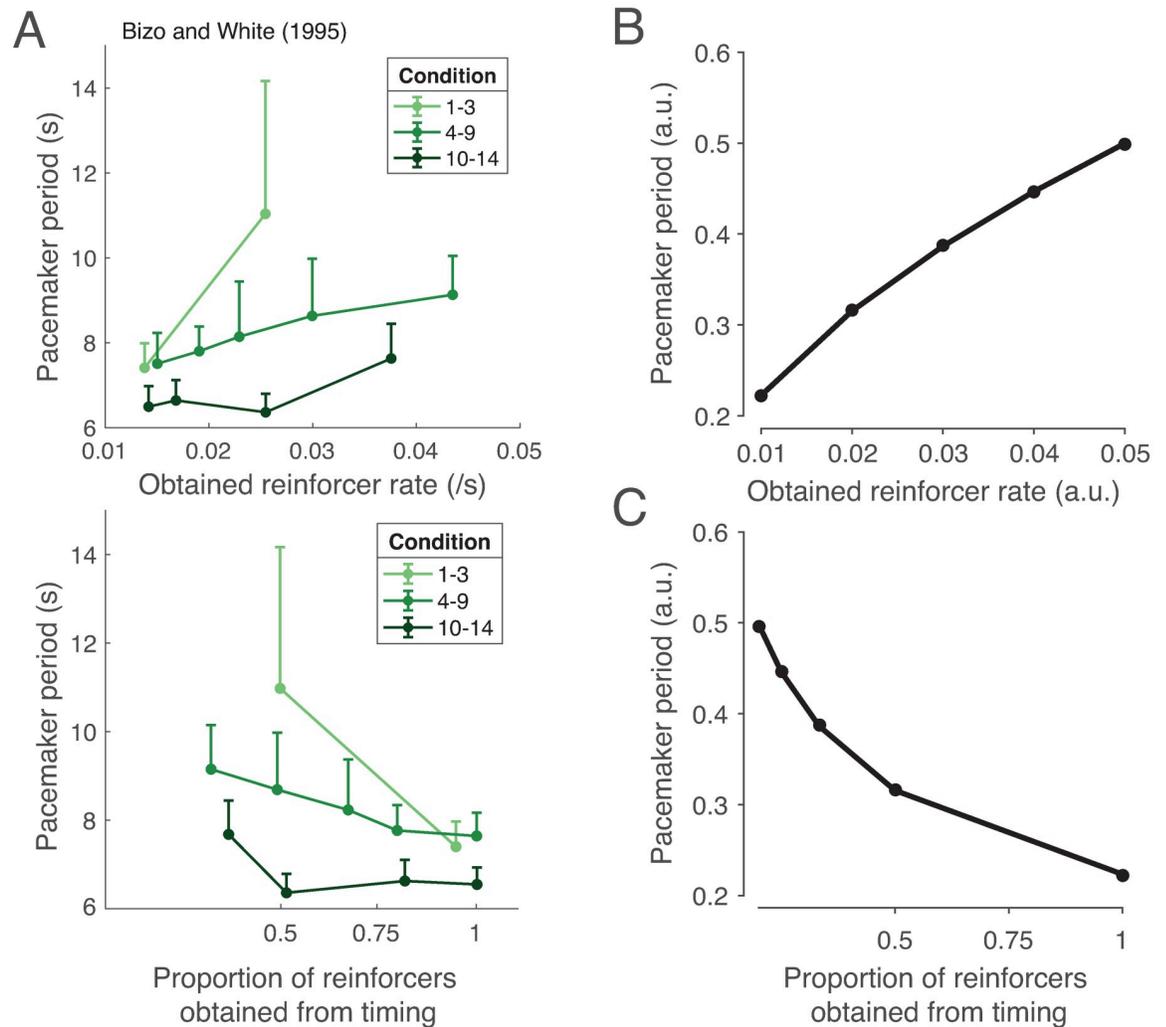


Fig 6. Rational inattention and controllability. (A) Controllability increases clock speed. Top panel: Bizo and White [134] have shown that when reinforcers are freely delivered, clock speed decreases (plotted here is the pacemaker period, or the inverse of clock speed, a parameter in their computational model which was fit to the empirical data). Bottom panel: On the other hand, when obtaining reinforcers is contingent on adequate timekeeping, clock speed increases. Light green, green, and dark green denote conditions in which free rewards were delivered during the intertrial interval, during both the intertrial interval and the trial, and during the trial only. We do not make a distinction among these conditions in our model. Figure adapted from [134]. (B, C) Our model recapitulates this effect: Under rational inattention, high average reward should only increase precision when this increase improves the ability to obtain rewards. (B) As free rewards increase and the added incentive of timing-contingent rewards decreases, clock speed will decrease. (C) On the other hand, as timing-contingent rewards increase, clock speed will increase. See [Methods](#) for simulation details.

<https://doi.org/10.1371/journal.pcbi.1008659.g006>

task, given the potential usefulness of observational learning for future performance. Additionally, in the real world, tasks overlap (a predator preparing to pounce uses the same visual information to identify its prey, assess its fitness, localize it well, and predict its next move), so reducing attention in a single subtask to zero without affecting the others is often not feasible.

Rational inattention also reconciles a longstanding question on temporal discounting and post-reward delays in animals: A large body of work has used intertemporal choice tasks to study the impact of delays on reward preferences in animals. In these studies, animals are given the choice between a small reward delivered soon and a large reward delivered later. Choosing the smaller (sooner) reward has been taken to indicate a discounting of future rewards, with the extent of discounting often taken to reflect qualities like impulsivity and

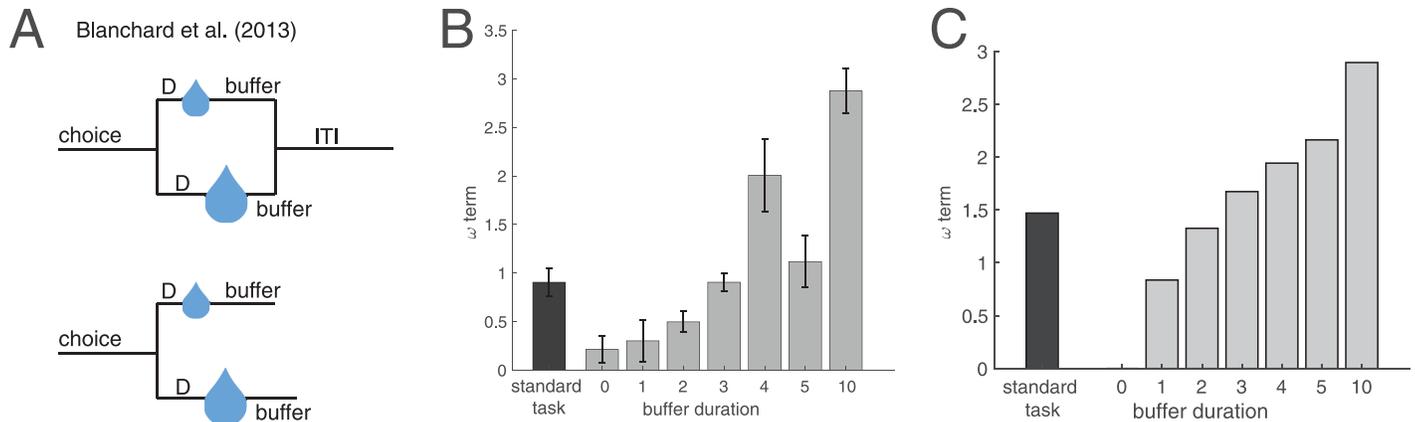


Fig 7. Rational inattention and post-reward delays. (A) Blanchard et al. [139] trained monkeys on an intertemporal choice task involving a small reward delivered soon or a large reward delivered later. Top: In the standard task, the total trial duration, or the sum of pre-reward delay ('D') and post-reward delay ('buffer'), was fixed to 6 seconds. The average buffer duration was 3 seconds. Bottom: In the constant buffer task, the post-reward delay was fixed regardless of choice, and was either 0, 1, 2, 3, 4, 5, or 10 seconds. ITI: intertrial interval. (B) Monkeys underestimate post-reward delays. By fitting monkey behavior to a model in which animals maximize long-term reward rate, the authors showed that the model fit value for the subjective estimate of post-reward delay (the ' ω term,' described in the [Methods](#)) is smaller than its true value (buffer duration). This relationship held for the standard task and across all buffer durations in the constant buffer task. For (A, B), figures adapted from [139]. (C) Our model recapitulates this effect: Under rational inattention, precision increases during the period when outcomes can be more strongly controlled (i.e., the pre-reward delay), and decreases otherwise (i.e., the post-reward delay, before the subsequent trial begins). This results in an underestimation of post-reward delays. See [Methods](#) for computational model and simulation details.

<https://doi.org/10.1371/journal.pcbi.1008659.g007>

self-control. A closer look, however, has suggested that animals' behaviors are consistent with an underestimation of time during the post-reward delay period, which is typically included after choosing the small reward in order to control for total trial durations and average reward rates [137–139]. Thus the apparent temporal discounting may simply arise from this underestimation of time during the post-reward delay, independent of reward value. For instance, Blanchard et al. [139] tested monkeys on an intertemporal choice task in which the post-reward delay was varied (Fig 7A). By motivating and fitting a model in which the animals maximized long-term reward rate, they showed that the animals systematically underestimated the post-reward delays (Fig 7B). The cause of this underestimation is still an open question [139]. However, rational inattention predicts this effect, as animals have more control over the outcome of the task before reward presentation than after it. Thus our framework predicts a slower clock during the post-reward delay, and behaviors consistent with an underestimation of time (Fig 7C).

This result makes the untested prediction that optogenetically stimulating DA neurons during the post-reward delay in tasks with fixed trial length should make animals less impulsive. Note here that the animal's behavior will appear to be impulsive with a *faster* clock in the *pre*-reward delay, or a *slower* clock in the *post*-reward delay. This is because the larger/earlier option has a longer pre-reward delay, but the smaller/sooner option has a longer post-reward delay. Thus a faster clock during the pre-reward delay will disproportionately increase the perceived total trial length of the larger/earlier option (higher impulsivity), whereas a faster clock during the post-reward delay will disproportionately increase the perceived total trial length of the smaller/sooner option (lower impulsivity).

It is important to note that the discounting-free model does not invalidate reward discounting in general. While the intertemporal choice task in animals involves training over many trials, discounting in humans typically involves mental simulations ("Do you prefer \$1 today or \$10 in one month?") that may involve neither an experienced pre-reward delay nor a post-reward delay. Nonetheless, humans systematically discount future rewards in these tasks [140].

Experimental predictions

The rational inattention framework makes a number of novel experimental predictions. Broadly, these can be classified based on any of three experimental manipulations (controllability, average reward, and DA level), two time courses (chronic and acute), two domains (reinforcement learning and interval timing), and two readouts (DA and behavior). Let us illustrate these with the testable example of controllability in reinforcement learning.

Consider a reinforcement learning task in which animals or humans are trained on a two-armed bandit, where arm *A* yields a small reward and arm *B* yields a large reward. In Experiment 1, subjects can sample the arms freely (high controllability), and the arms are denoted by A_1 and B_1 . In Experiment 2, subjects are merely observers, so the arms are sampled for them (no controllability), and the arms are denoted by A_2 and B_2 . Arms A_1 and A_2 yield the same reward, but each is accompanied by a distinct stimulus (e.g., sound); similarly for arms B_1 and B_2 . After training on Experiments 1 and 2 separately, subjects are tested on a two-armed bandit consisting of arms A_1 and A_2 or arms B_1 and B_2 (each with its accompanying stimulus, so that the arms are distinguishable). The rational inattention framework predicts that, because of the central tendency, A_2 will be more likely to be selected than A_1 , whereas B_1 will be more likely to be selected than B_2 (Fig 8A). This effect cannot be explained by a preference for one type of controllability over the other: For B, the option trained under a *more* controllable context is preferred, whereas for A, the option trained under a *less* controllable context is preferred. Similarly, the effect cannot be explained by assuming better learning in one experiment over the other. Finally, this setup can control for any differences in sampling between the two experiments, as the choices in Experiment 2, which are made by a computer, can exactly match those made by the subject in Experiment 1.

Analogs of this experiment can be performed while varying average reward or DA levels instead of controllability. In these cases, the probability of selecting each action during training can further be compared across experiments: The probability difference in choosing each arm will be greater in Experiment 1 (high average reward or high DA) than in Experiment 2 (low average reward or low DA) due to the central tendency. This is similar to the experiment by Cinotti et al. [30] discussed previously, and can be interpreted as the animal exploiting under high average reward and high DA, and exploring under low average reward and low DA. (Note here that differences in sampling frequency for each arm are not controlled for.)

A second experiment can test value estimates by directly measuring *phasic* DA responses, which putatively report 'reward prediction errors' (RPEs), or the difference between received and expected rewards [1–3, 21, 22]. Consider then a task in which animals or humans are again trained on a two-armed bandit, either with (Experiment 1) or without (Experiment 2) controllability, but where one arm is stochastic. In particular, let arm *A* yield a reward that is either small or large, with equal probability. After some learning, when arm *A* is chosen, the large reward will elicit a phasic DA burst (received reward is larger than expected), whereas the small reward will elicit a dip (received reward is smaller than expected). This will occur in both Experiments 1 and 2. However, the phasic burst following A_1 when a large reward is received will be greater than that following A_2 (we use the same subscript notation as above). On the other hand, the dip following A_1 when a small reward is received will be deeper than that following A_2 , again due to the central tendency (Fig 8B). As above, this effect cannot be explained by a preference for one type of controllability or preferential learning in one experiment over the other. Also as above, analogs of this experiment can be performed by replacing controllability with average reward or tonic DA levels.

Third, using the same principles, we can design an experiment to distinguish between the gain control hypothesis of tonic DA and the rational inattention framework. Recall that, to

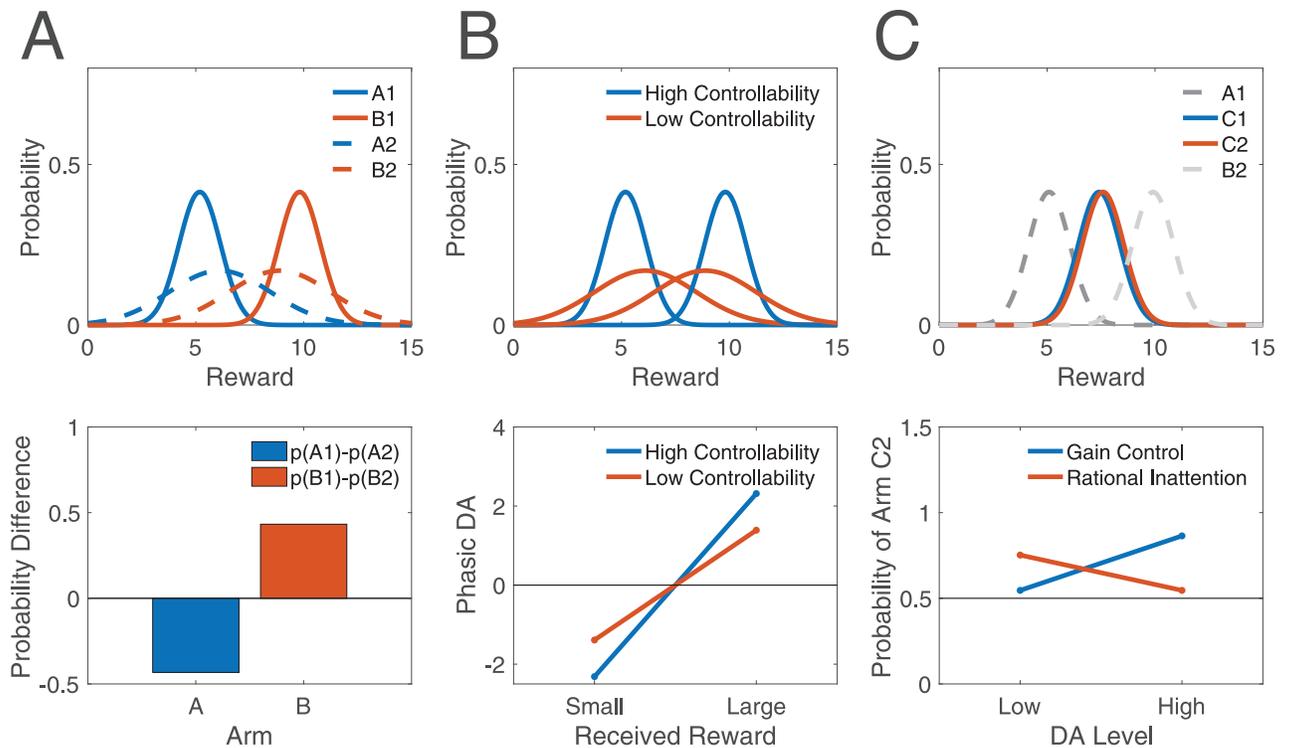


Fig 8. Experimental predictions of rational inattention. (A) Top panel: The difference in value estimates for two arms is higher under high controllability (solid curves) than low controllability (dashed curves). Bottom panel: After learning under each condition, the two arms yielding small rewards are compared. Rational inattention predicts that the arm trained under low controllability will be selected more often. On the other hand, when the two arms yielding large rewards are compared, that trained under high controllability will be selected more often. $p(A1)$: probability of selecting arm A1; similarly for B1, A2, and B2. (B) Top panel: When a single arm yields a small or large reward with equal probability, the estimated deviation of actual outcomes from the mean reward will have larger magnitude under high-controllability learning than under low-controllability learning. Bottom panel: Thus the small reward will elicit a more negative phasic DA response, and the large reward will elicit a more positive phasic DA response, under high controllability than under low controllability. (C) Top panel: Arms C1 and C2 are identical, but C1 is trained with an arm yielding a smaller reward (A1), and C2 is trained with an arm yielding a larger reward (B2). The estimates for the two identical arms will be on opposite sides of their true value due to the central tendency. Bottom panel: After training, arms C1 and C2 are combined into a new two-armed bandit task, occurring under either high or low DA. The gain control hypothesis of DA predicts that the difference in their estimates will be amplified under high DA, thus making selection of C2 more likely than under low DA. On the other hand, rational inattention predicts that the central tendency will be reduced under high DA, which, in this task, will cause the two estimates to migrate closer to their true reward value (and therefore, to each other), in turn making selection of C2 less likely than under low DA. See [Methods](#) for simulation details.

<https://doi.org/10.1371/journal.pcbi.1008659.g008>

explain tonic DA's role in performance (as opposed to its role in learning), a number of authors have posited that tonic DA implements some form of 'gain control' on action values, whereby differences in action values between different options are amplified by high tonic DA levels, which promotes exploitation of the most rewarding option. The rational inattention framework subsumed this idea by showing that controlling precision mimics gain control, at least in simple cases ([Methods](#)). Let us then design an experiment to distinguish between these alternatives: Consider a two-armed bandit task of small and medium reward (Experiment 1) and medium and large reward (Experiment 2). After training on these pairs independently, the two identical (or nearly identical) medium rewards—one from each experiment—are combined into a new pair. The subject must make selections in this new two-armed bandit task, under either high or low DA. Rational inattention predicts that, because of the central tendency, the medium reward from Experiment 2 will be selected more often under low DA, with a smaller (or absent) effect under high DA. But this is the exact opposite prediction of the gain control hypothesis, whereby any differences in selection should be amplified under high DA

(Fig 8C). As a proof of concept, this result may begin to explain why some studies have found exploration to be enhanced under high DA [31], which thus far has been viewed as incompatible with the seemingly competing literature discussed above.

Discussion

Questions on the roles of tonic DA abound. While two theories have put forth compelling arguments attributing tonic DA to either precision or average reward, it has remained unclear conceptually whether and how these two quantities are related to each other. Furthermore, within each domain, questions arise. Under the precision viewpoint, why would fluctuations in tonic DA separately influence both true and estimated precision, a seemingly suboptimal strategy when encoding and decoding are temporally decoupled? In reinforcement learning models, how and why does tonic DA implement gain control on action values during performance, and how does this relate to its role in learning? Rational inattention resolves these questions: By reporting the single quantity of context-specific average reward, DA first determines the precision with which encoding occurs, and second, faithfully relays the precision used during encoding to the decoding stage. This view unifies the two theories, while simultaneously endogenizing the ‘gain control’ function of DA and safeguarding against the suboptimality of precision miscalibration. Beyond DA, this framework takes an additional step toward integrating theories of reinforcement learning and interval timing.

In reinforcement learning, the rational inattention framework predicts that learning from positive and negative feedback is enhanced under high and low DA, respectively: Because DA signals average reward, an animal under high DA expects high rewards in these tasks. Thus under high DA, positive feedback is expected, and therefore more readily learned, whereas under low DA, negative feedback is more readily learned. Second, under rational inattention, high DA suppresses the contribution of context to the final estimate. Thus when two reward magnitudes are learned and subsequently compared, this suppression of interfering signals increases the difference between the estimated magnitudes. This in turn tips the exploration-exploitation balance toward exploitation of the higher reward.

In interval timing, rational inattention predicts that both high DA levels and high average reward result in a faster internal clock: Constrained by time cell rescaling, increases in precision occur through a compression of subjective time against objective time, and thus lead to a faster internal clock. We take these scalable time cells to constitute the internal timing mechanism [126, 127], which may be well-adapted to rapid learning of the temporal properties of an environment [141]. Similarly, low DA levels and low average reward lead to a slower clock. Low DA simultaneously increases the relative contribution of context (other signals) to the temporal estimate, which exaggerates the central tendency effect. Finally, rational inattention also predicts that the modulation of precision by average reward should only apply when obtaining the rewards is contingent on the agent’s performance. When rewards are freely delivered, precision should not increase, as it would be wasteful to spend cognitive resources on the present task. Faced with enough rewards that are not dependent on performance, the clock should instead slow down. The effect of controllability also implies that animals will underestimate the duration in a task that follows reward delivery but precedes the end of each trial, a long-known but poorly understood phenomenon of post-reward delays.

Rational inattention captures empirical phenomena that cannot be captured by either the average reward theory or the Bayesian theory alone (Table 1). While the average reward theory succeeds in explaining the effect of learning from positive vs. negative feedback, it fails in capturing all other presented phenomena. The Bayesian model is more successful, and can be extended to capture the exploration-exploitation findings in reinforcement learning as well as

Table 1. Summary of predicted phenomena by theory. The average reward theory is successful when controllability, DA, or average reward is manipulated, and DA or the reward estimate is measured. The Bayesian theory is successful when DA is manipulated, and behavior is measured. The rational inattention framework is successful with any combination of these variables.

		Average reward theory	Bayesian theory	Rational inattention
Reinforcement learning	DA and learning from positive vs negative feedback	✓	✗	✓
	DA and exploitation	✗	✓	✓
Interval timing	DA and the central tendency	✗	✓	✓
	DA and clock speed	✗	✓	✓
	Average reward and clock speed	✗	✗	✓
Controllability	Clock speed	✗	✗	✓
	Post-reward delays	✗	✗	✓

<https://doi.org/10.1371/journal.pcbi.1008659.t001>

the DA manipulations in interval timing (central tendency and clock speed), but fails in capturing the effect of learning. Both theories fail in predicting the effect of average reward on the speed of the internal clock, the effect of controllability, and post-reward delays. In other words, the individual theories fail when the experimental manipulations and measured outcomes correspond to variables upstream and downstream of DA, respectively (i.e., nodes above and below the green arrow in Fig 2, inclusive). When both are upstream of DA (thus involving controllability, average reward, and DA), the average reward theory is successful; when both are downstream of DA (thus involving DA and precision), the Bayesian theory is successful. As an example, consider the effect of controllability on the speed of the internal clock. The average reward theory predicts that animals should act more vigorously and with higher response rates when in high average-reward environments, which applies to timing-contingent and timing-noncontingent rewards alike. However, without some additional assumption, the theory does not predict that clock speed should increase in the first case and decrease in the second. Similarly, while the Bayesian theory can accommodate changes in clock speed, it lacks a theoretical foundation for why controllability should influence the clock (or even, as an intermediate step, the DA level).

The framework we have presented is related to the hypothesis that DA in some tasks mediates the ‘cost of control.’ Notably, recent work by Manohar et al. [142, 143] has shown that subjects break the speed-accuracy trade-off in a saccadic selection task when motivated by reward, simultaneously increasing their speed and improving their accuracy. The authors propose a cost of control, which can be overcome with rewards, an effect they hypothesize is mediated by DA. While our work generalizes this result at the normative and algorithmic levels, the question of how the precision-cost trade-off is implemented neurobiologically remains to be determined in future work. It should be emphasized here that our framework removes the confounding speed-accuracy trade-off element, because in the tasks we model, agents cannot earn more reward by responding more quickly. For instance, Otto and Daw [144] have found that, when subjects are given a set amount of time to sequentially collect as many rewards as they want, they show higher error rates. This makes sense from a normative perspective: Spending too much time on one trial forces subjects to forgo future rewards. So under high reward rates, subjects should act faster, even if it means a higher error rate, because the total number of accumulated rewards will go up.

It is important to note that the story of DA and performance is not as linear as we have assumed so far. Rather, a significant body of work has shown that when very high DA levels are reached, an inverse U-shaped relationship between DA and performance emerges [145–150]. Rational inattention, as we have presented it, predicts that precision, and therefore performance, simply increase with DA. However, it is reasonable to assume that the encoding

machinery has fundamental limits to its precision, so that arbitrary increases in DA may wildly increase estimated precision without an accompanying and equal increase in true precision. This will lead to precision miscalibration, and increases in DA over this range will now worsen the miscalibration and thus worsen performance. In [S5 Appendix](#), we derive this result analytically, and show that when true precision is bounded, the relationship between DA and performance takes the shape of an inverted U. This effect may also explain a number of apparent inconsistencies in the experimental literature. Notably, Beeler et al. [31] have found that chronically hyperdopaminergic mice were more willing to select high-cost levers in a two-armed bandit task than wild-type mice. Computational modeling of this experiment suggested that DA promotes exploration (rather than exploitation), which, as the authors note, may be explained by the U-shaped effect of DA. Similarly, behavior consistent with a *slower*, rather than faster, clock has been reported with optogenetic stimulation of midbrain DA neurons in mice [151] (but see [152, 153]), as well as in Parkinson's patients who were off medication during training but on medication during testing on a separate day [40]. Whether these seemingly inconsistent findings are owed to the U-shaped effect of DA and its manipulation at non-physiological levels remains to be examined.

In restricting our analysis to the effects of tonic DA in reinforcement learning and interval timing, we have disregarded a wide array of experimental findings on DA. For instance, DA exhibits a distinct phenomenon of 'ramping' over the course of a single trial in a number of reinforcement learning tasks, such as goal-directed spatial navigation [154], bandit tasks [32], and timing of movement initiation [152], but not in others, such as classical conditioning tasks [1, 155–157]. These ramps, which occur on the timescale of seconds, are primarily observed during operant tasks, with a rapid return to baseline after task completion (e.g., during the post-reward delay). A natural question, then, is whether this differential in the average DA level before vs. after task completion mediates the effects of controllability on clock speed. Some authors have indeed interpreted these ramps as a 'quasi-tonic' signal [158], while others have argued in favor of an RPE interpretation of ramps, similar to that of phasic DA signals [153, 159–161]. DA has also been implicated in mediating spatiotemporal credit assignment in reinforcement learning tasks [162], and DA's roles in working memory [163–166], spontaneous movement [167, 168], impulsivity [169–174], creativity [175, 176], and other domains have been the subject of great interest as well.

Second, seeking a computational theory of tonic DA necessarily introduced a number of simplifications. For instance, DA's effects vary by receptor subtype: In the basal ganglia, which implements reinforcement learning models, the neurons primarily expressing either D1 or D2 receptors largely segregate anatomically into two separate pathways (the 'direct' and 'indirect' pathways, respectively, which later converge) [177] and seem to serve opposite purposes [178, 179]. DA bursts primarily potentiate D1 synaptic weights and depress D2 synaptic weights, and vice versa for DA dips [180]. Furthermore, the opposing effects of phasic DA on D1 and D2 receptors seems to extend to tonic DA as well [27, 181, 182]. On the other hand, based on pharmacological studies using D1 and D2 antagonists, the interval timing effects of DA seem to be primarily D2-mediated [113, 183], although recent work has highlighted a role for D1 as well [56, 113, 183–185]. While computational theories should transcend specific implementations [186], a computationally complete picture of DA will likely need to account for receptor heterogeneity [4, 27]. DA's effects similarly vary by projection site (with broad projections across cortical and subcortical regions [187, 188]) and enzymatic activity [6, 189], and the predictions of the rational inattention framework cut across this diversity. For example, consider tonic DA's role in the exploration-exploitation trade-off. Recent work has shown that DA's action in prefrontal cortex modulates an exploration strategy referred to as 'directed exploration,' in which uncertainty about a reward source confers a value bonus, thus making the

source more likely to be sampled. Striatal DA, on the other hand, has been linked to ‘random exploration,’ in which agents modify the stochasticity of their choices according to the total uncertainty in the environment [6, 190]. How these empirical findings may fit into a broader theory of tonic DA will be the subject of future work.

Third, in adopting the rational inattention framework, we have defined the ‘cognitive cost’ to be the cost of reducing the uncertainty in the world. This is far from the only cognitive cost an animal must pay in a task, and indeed different costs may outweigh others depending on the task. For a theory of tonic DA, our attentional cost function was informed by the empirical evidence linking DA to precision. However, a more complete optimization problem will need to incorporate capacity costs [191], computation costs [192], interference costs [193], metabolic costs [194], and others [195]. How these different factors influence the optimization problem, and whether and how they interact with DA, remain to be examined.

Finally, this analysis does not preclude other factors from influencing the precision of encoding or the final estimate of decoding. Using reinforcement learning as an example, the volatility of an environment [196] and the stochasticity of reward sources [197] should affect the learning rate, but it is not clear that these quantities are reflected by the DA signal and rather are likely under direct cortical control [196, 198, 199] (but see [200, 201]). We have presented here the base case where, holding all else constant, average reward controls encoding and decoding in a very straightforward way. In more realistic environments where all else is not held constant, average reward will be one of a number of factors influencing encoding and subsequent decoding.

At first glance, the functions of DA seem to vary across processing stage and modality. We have shown how seemingly unrelated behaviors—such as modulation of the speed of an internal clock and learning from positive feedback—can be traced back to similar computations under the unifying principle of rational inattention.

Methods

Precision and gain modulation of value in reinforcement learning

In the [Results](#), we argue that manipulating estimated precision has the apparent effect of controlling the gain of action values. Here, we describe this prediction concretely.

Standard models of action selection posit that the probability of selecting action A_i with estimated value $\hat{\mu}_i$ follows a softmax function [105, 106]:

$$p(A_i) = \frac{e^{\beta \hat{\mu}_i}}{\sum_j e^{\beta \hat{\mu}_j}}, \quad (13)$$

where β is referred to as the inverse temperature parameter. Recent influential models have argued that DA modulates the gain of the values $\hat{\mu}_i$, possibly by controlling the inverse temperature parameter β [27, 30, 202]. For simplicity, we examine the case of two actions, A_l and A_s , associated with large and small reward, respectively. [Eq 13](#) then reduces to

$$p(A_l) = \frac{1}{1 + e^{-\beta(\hat{\mu}_l - \hat{\mu}_s)}}. \quad (14)$$

Importantly, the probability of selecting the large reward (exploitation) depends on the difference between the reward magnitudes and not on the absolute magnitudes themselves. As the quantity $\beta(\hat{\mu}_l - \hat{\mu}_s)$ decreases, $p(A_l)$ decreases. Hence, any manipulation that results in a decrease in the estimated difference will encourage exploration over exploitation. Gain control is conventionally viewed as acting through the inverse temperature parameter β , which serves to amplify or reduce the influence of the difference $(\hat{\mu}_l - \hat{\mu}_s)$ on the animal’s behavior. In this

simple case, modifying β can be thought of as modifying choice stochasticity (see [203, 204] for more precise formulations). However, the same effect can be achieved by amplifying or reducing the estimated difference ($\hat{\mu}_i - \hat{\mu}_s$) directly. Under Bayes' rule, manipulating the likelihood precisions modulates the resulting difference in posterior means, thus implementing gain control on action values.

Computational model of post-reward delay

We consider now models in which animals make their choices by simply maximizing the average reward over the entire task [205, 206]. In particular, Blanchard et al. [139] propose a discounting-free model of intertemporal choice, in which the animal maximizes non-discounted average reward (total reward divided by total trial duration). Thus the value v_c of each choice c can be written as

$$v_c = \frac{r_c}{D + w}, \quad (15)$$

where r_c is the actual reward for choice c , D is the pre-reward delay, and w is the estimated post-reward delay, which is a free parameter fit by behavior. Importantly, this model is mathematically translatable to hyperbolic discounting models, which also have a single parameter (in this case, the temporal discount factor k):

$$v_c = \frac{r_c}{1 + kD}. \quad (16)$$

Thus, as the authors note, the data is fit by both models equally well, and cannot be used to arbitrate between them. (The authors argue against the discounting model elsewhere in their study, with Eq 15 simply serving to examine the magnitude of w against the true post-reward delay.)

Simulation details

Reinforcement learning. For the exploration-exploitation trade-off, we have chosen $\kappa = 0.1$, and reward magnitude of 1. Baseline DA was set to 0.9, and prior precision to 10. In all conditions, average reward was set to the DA level. Action selection was implemented using the softmax function in Eq 13, with $\beta = 4$. Parameter tuning: The qualitative results hold for any choice of $\frac{\beta}{\kappa} > 0$, $\lambda_0 > 0$, $DA > 0$, and $\beta > 0$, such that $\frac{2\beta}{\kappa} > \lambda_0$ (after Eq 10) (Fig 4B). For learning from positive and negative feedback, we have chosen $\kappa = 0.1$, and reward magnitudes of 1 and 0 to reflect reward and punishment, respectively. To model the effect of average reward, reported by DA, on prior expectations in a new context, we set the prior mean to equal the DA level. (Note that, in general, the prior mean may be a combination of the average reward and recently experienced stimuli, weighted by their precisions.) Prior precision was arbitrarily set to 5. Accuracy was operationally defined as the area under the posterior distribution closer to the correct outcome (1 for rewards, 0 for punishments; i.e., the area to the right of and to the left of 0.5, respectively). Relative accuracy is the difference between these two areas. Parameter tuning: The qualitative results hold for any choice of $\frac{\beta}{\kappa} > 0$, μ , $\lambda_0 > 0$, and $DA > 0$ (Fig 4D).

Interval timing. For both experiments, our objective-subjective mapping is $m = \eta \log(\mu + 1)$ with subjective precision $l = 1$ to capture Weber's law (see S1 Appendix for a derivation). Average reward is equal to the DA level. For the central tendency, $\kappa = 0.05s^{-1}$, and $DA = 0.2$ and 1 for low and high DA levels, respectively. By visual comparison with Shi et al. [41], we approximated the prior standard deviation to be 2.5 times smaller than the difference between

presented durations. Prior precision is the inverse square of the prior standard deviation. Parameter tuning: The qualitative results hold for any choice of $\frac{R}{\kappa} > 0$, $l > 0$, $\lambda_0 > 0$, and $DA > 0$, such that $\frac{2R}{\kappa} > \lambda_0$ (after Eq 10) (Fig 5B). For the speed of the internal clock, $\kappa = 0.1s^{-1}$, $DA = 0.8, 1$, and 1.2 for the low, baseline, and high DA conditions, respectively. Parameter tuning: The qualitative results hold for any choice of $\frac{R}{\kappa} > 0$, $l > 0$, and $DA > 0$ (Fig 5D).

Controllability. For both experiments, we set average reward to be equal to the DA level. Rate of timing-contingent rewards was set to 0.01, and $\kappa = 0.1s^{-1}$. Note here that the reward incentive R formally refers to the subjective value of the reward, rather than its objective value. We have generally treated these quantities as interchangeable, as, under normal circumstances, they are monotonically related and therefore affect precision in the same qualitative way (Eq 10). However, one can specifically disentangle these two by manipulating satiety or baseline reward availability, as in Bizo and White [134]. Indeed, the incentive to accumulate rewards in satiated states is very low; the incentive to accumulate the same amount of reward but at near starvation is extremely high. To capture this distinction, we set R to be a decreasing function of baseline reward availability (more precisely, we set them to be inversely related). Note that this effect can also be captured by taking subjective reward to be a concave function of objective reward, following convention [207, 208]. This reflects the idea that the added benefit of an extra unit of reward decreases as more rewards are accumulated. Parameter tuning: The qualitative results hold for any choice of $\frac{R}{\kappa} > 0$ and $DA > 0$, where R is a decreasing function of free rewards (Fig 6B and 6C). For post-reward delays, we arbitrarily set average reward to be 1, with a baseline of 0.7 for the post-reward delay. Parameter tuning: The qualitative results hold for any choice of $\frac{R}{\kappa} > 0$ and $DA > 0$, where R is larger for the pre-reward delay than for the post-reward delay (Fig 7C).

Experimental predictions. Small, medium, and large rewards had magnitude 5, 7.5, and 10, respectively. Precision under high and low controllability was set to 1 and 0.1, respectively. Action selection was implemented using the softmax function in Eq 13, with $\beta = 1$ (Fig 8). RPEs were computed as the difference between received reward (5 or 10) and the expected reward for the arm (7.5) (Fig 8B). To simulate gain control, we set the inverse temperature parameter to $\beta = 1$ for low DA and $\beta = 10$ for high DA (Fig 8C). Parameter tuning: For all predictions, the qualitative results hold for any choice of $\frac{R}{\kappa} > 0$, μ , $\lambda_0 > 0$, $DA > 0$, and $\beta > 0$, such that $\frac{2R}{\kappa} > \lambda_0$ (after Eq 10).

Supporting information

S1 Appendix. The scalar property as a consequence of rational inattention.

(PDF)

S2 Appendix. Time cell rescaling.

(PDF)

S3 Appendix. Interval timing in discrimination tasks.

(PDF)

S4 Appendix. Interval timing under average reward manipulations.

(PDF)

S5 Appendix. Inverse U-shaped relationship between DA and performance.

(PDF)

Acknowledgments

The authors are grateful to Rahul Bhui for comments on the manuscript.

Author Contributions

Conceptualization: John G. Mikhael, Samuel J. Gershman.

Formal analysis: John G. Mikhael, Samuel J. Gershman.

Funding acquisition: Samuel J. Gershman.

Investigation: John G. Mikhael.

Methodology: John G. Mikhael.

Resources: Samuel J. Gershman.

Software: John G. Mikhael, Lucy Lai.

Supervision: Samuel J. Gershman.

Writing – original draft: John G. Mikhael.

Writing – review & editing: John G. Mikhael, Lucy Lai, Samuel J. Gershman.

References

1. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science*. 1997; 275(5306):1593–1599. <https://doi.org/10.1126/science.275.5306.1593> PMID: 9054347
2. Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH. A causal link between prediction errors, dopamine neurons and learning. *Nature neuroscience*. 2013; 16(7):966–973. <https://doi.org/10.1038/nn.3413> PMID: 23708143
3. Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N. Arithmetic and local circuitry underlying dopamine prediction errors. *Nature*. 2015; 525:243–246. <https://doi.org/10.1038/nature14855> PMID: 26322583
4. Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science*. 2004; 306:1940–1943. <https://doi.org/10.1126/science.1102941> PMID: 15528409
5. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*. 2007; 104(41):16311–16316. <https://doi.org/10.1073/pnas.0706111104> PMID: 17913879
6. Frank MJ, Doll BB, Oas-Terpstra J, Moreno F. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature neuroscience*. 2009; 12(8):1062. <https://doi.org/10.1038/nn.2342> PMID: 19620978
7. Howard CD, Li H, Geddes CE, Jin X. Dynamic nigrostriatal dopamine biases action selection. *Neuron*. 2017; 93(6):1436–1450. <https://doi.org/10.1016/j.neuron.2017.02.029> PMID: 28285820
8. Geddes CE, Li H, Jin X. Optogenetic Editing Reveals the Hierarchical Organization of Learned Action Sequences. *Cell*. 2018; 174(1):32–43. <https://doi.org/10.1016/j.cell.2018.06.012> PMID: 29958111
9. Nonomura S, Nishizawa K, Sakai Y, Kawaguchi Y, Kato S, Uchigashima M, et al. Monitoring and updating of action selection for goal-directed behavior through the striatal direct and indirect pathways. *Neuron*. 2018; 99(6):1302–1314. <https://doi.org/10.1016/j.neuron.2018.08.002> PMID: 30146299
10. Wang DV, Tsien JZ. Conjunctive processing of locomotor signals by the ventral tegmental area neuronal population. *PLoS One*. 2011; 6(1):e16528. <https://doi.org/10.1371/journal.pone.0016528> PMID: 21304590
11. Yin HH. Action, time and the basal ganglia. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2014; 369(1637):20120473. <https://doi.org/10.1098/rstb.2012.0473> PMID: 24446506
12. Barter JW, Li S, Lu D, Bartholomew RA, Rossi MA, Shoemaker CT, et al. Beyond reward prediction errors: the role of dopamine in movement kinematics. *Frontiers in integrative neuroscience*. 2015; 9:39. <https://doi.org/10.3389/fnint.2015.00039> PMID: 26074791
13. Noudoost B, Moore T. Control of visual cortical signals by prefrontal dopamine. *Nature*. 2011; 474(7351):372. <https://doi.org/10.1038/nature09995> PMID: 21572439

14. Stalter M, Westendorff S, Nieder A. Dopamine Gates Visual Signals in Monkey Prefrontal Cortex Neurons. *Cell Reports*. 2020; 30(1):164–172. <https://doi.org/10.1016/j.celrep.2019.11.082> PMID: 31914383
15. Maricq AV, Roberts S, Church RM. Methamphetamine and time estimation. *Journal of Experimental Psychology: Animal Behavior Processes*. 1981; 7(1):18. PMID: 7229573
16. Maricq AV, Church RM. The differential effects of haloperidol and methamphetamine on time estimation in the rat. *Psychopharmacology*. 1983; 79(1):10–15. <https://doi.org/10.1007/BF00433008> PMID: 6403957
17. Lake JI, Meck WH. Differential effects of amphetamine and haloperidol on temporal reproduction: dopaminergic regulation of attention and clock speed. *Neuropsychologia*. 2013; 51(2):284–292. <https://doi.org/10.1016/j.neuropsychologia.2012.09.014> PMID: 22982605
18. Soares S, Atallah BV, Paton JJ. Midbrain dopamine neurons control judgment of time. *Science*. 2016; 354(6317):1273–1277. <https://doi.org/10.1126/science.aah5234> PMID: 27940870
19. Nieoullon A. Dopamine and the regulation of cognition and attention. *Progress in neurobiology*. 2002; 67(1):53–83. [https://doi.org/10.1016/S0301-0082\(02\)00011-4](https://doi.org/10.1016/S0301-0082(02)00011-4) PMID: 12126656
20. Grace AA. Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience*. 1991; 41(1):1–24. [https://doi.org/10.1016/0306-4522\(91\)90196-U](https://doi.org/10.1016/0306-4522(91)90196-U) PMID: 1676137
21. Niv Y, Schoenbaum G. Dialogues on prediction errors. *Trends in cognitive sciences*. 2008; 12(7):265–272. <https://doi.org/10.1016/j.tics.2008.03.006> PMID: 18567531
22. Glimcher PW. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*. 2011; 108(Supplement 3):15647–15654. <https://doi.org/10.1073/pnas.1014269108> PMID: 21389268
23. 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. *Journal of Neuroscience*. 2009; 29(5):1538–1543. <https://doi.org/10.1523/JNEUROSCI.4467-08.2009> PMID: 19193900
24. Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P, et al. Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. *Brain*. 2012; 135(6):1871–1883. <https://doi.org/10.1093/brain/aws083> PMID: 22508958
25. Smittenaar P, Chase H, Aarts E, Nusslein B, Bloem B, Cools R. Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection—learning or performance? *European Journal of Neuroscience*. 2012; 35(7):1144–1151. <https://doi.org/10.1111/j.1460-9568.2012.08043.x> PMID: 22487043
26. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*. 2006; 442(7106):1042. <https://doi.org/10.1038/nature05051> PMID: 16929307
27. Collins AG, Frank MJ. Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological review*. 2014; 121(3):337. <https://doi.org/10.1037/a0037015> PMID: 25090423
28. Eisenegger C, Naef M, Linssen A, Clark L, Gandamaneni PK, Müller U, et al. Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology*. 2014; 39(10):2366. <https://doi.org/10.1038/npp.2014.84> PMID: 24713613
29. Lee E, Seo M, Dal Monte O, Averbeck BB. Injection of a dopamine type 2 receptor antagonist into the dorsal striatum disrupts choices driven by previous outcomes, but not perceptual inference. *Journal of Neuroscience*. 2015; 35(16):6298–6306. <https://doi.org/10.1523/JNEUROSCI.4561-14.2015> PMID: 25904783
30. Cinotti F, Fresno V, Akil N, Coutureau E, Girard B, Marchand AR, et al. Dopamine blockade impairs the exploration-exploitation trade-off in rats. *Scientific reports*. 2019; 9(1):6770. <https://doi.org/10.1038/s41598-019-43245-z> PMID: 31043685
31. Beeler JA, Daw ND, Frazier CR, Zhuang X. Tonic dopamine modulates exploitation of reward learning. *Frontiers in behavioral neuroscience*. 2010; 4:170. <https://doi.org/10.3389/fnbeh.2010.00170> PMID: 21120145
32. Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, et al. Mesolimbic dopamine signals the value of work. *Nature Neuroscience*. 2016; 19:117–126. <https://doi.org/10.1038/nn.4173> PMID: 26595651
33. Salamone JD, Cousins MS, Bucher S. Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit

- procedure. *Behavioural brain research*. 1994; 65(2):221–229. [https://doi.org/10.1016/0166-4328\(94\)90108-2](https://doi.org/10.1016/0166-4328(94)90108-2) PMID: 7718155
34. Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*. 2007; 191(3):507–520. <https://doi.org/10.1007/s00213-006-0502-4> PMID: 17031711
 35. Chong TTJ, Bonnelle V, Manohar S, Veromann KR, Muhammed K, Tofaris GK, et al. Dopamine enhances willingness to exert effort for reward in Parkinson's disease. *cortex*. 2015; 69:40–46. <https://doi.org/10.1016/j.cortex.2015.04.003> PMID: 25967086
 36. Beeler JA, Frank MJ, McDaid J, Alexander E, Turkson S, Bernandez MS, et al. A role for dopamine-mediated learning in the pathophysiology and treatment of Parkinson's disease. *Cell reports*. 2012; 2(6):1747–1761. <https://doi.org/10.1016/j.celrep.2012.11.014> PMID: 23246005
 37. Meck WH. Affinity for the dopamine D2 receptor predicts neuroleptic potency in decreasing the speed of an internal clock. *Pharmacology Biochemistry and Behavior*. 1986; 25(6):1185–1189. [https://doi.org/10.1016/0091-3057\(86\)90109-7](https://doi.org/10.1016/0091-3057(86)90109-7) PMID: 2880350
 38. 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. *Neurobiology of learning and memory*. 2007; 88(2):149–159. <https://doi.org/10.1016/j.nlm.2007.04.005> PMID: 17513138
 39. Malapani C, Rakitin B, Levy R, Meck WH, Deweer B, Dubois B, et al. Coupled temporal memories in Parkinson's disease: a dopamine-related dysfunction. *Journal of Cognitive Neuroscience*. 1998; 10(3):316–331. <https://doi.org/10.1162/089892998562762> PMID: 9869707
 40. Malapani C, Deweer B, Gibbon J. Separating storage from retrieval dysfunction of temporal memory in Parkinson's disease. *Journal of Cognitive Neuroscience*. 2002; 14(2):311–322. <https://doi.org/10.1162/089892902317236920> PMID: 11970794
 41. Shi Z, Church RM, Meck WH. Bayesian optimization of time perception. *Trends in Cognitive Sciences*. 2013; 17(11):556–564. <https://doi.org/10.1016/j.tics.2013.09.009> PMID: 24139486
 42. Jazayeri M, Shadlen MN. Temporal context calibrates interval timing. *Nature neuroscience*. 2010; 13(8):1020. <https://doi.org/10.1038/nn.2590> PMID: 20581842
 43. Acerbi L, Wolpert DM, Vijayakumar S. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS computational biology*. 2012; 8(11):e1002771. <https://doi.org/10.1371/journal.pcbi.1002771> PMID: 23209386
 44. Bausenhardt KM, Dyjas O, Ulrich R. Temporal reproductions are influenced by an internal reference: Explaining the Vierordt effect. *Acta Psychologica*. 2014; 147:60–67. <https://doi.org/10.1016/j.actpsy.2013.06.011> PMID: 23896562
 45. Mayer KM, Di Luca M, Ernst MO. Duration perception in crossmodally-defined intervals. *Acta psychologica*. 2014; 147:2–9. <https://doi.org/10.1016/j.actpsy.2013.07.009> PMID: 23953664
 46. Roach NW, McGraw PV, Whitaker DJ, Heron J. Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences*. 2017; 114(2):412–417. <https://doi.org/10.1073/pnas.1610706114> PMID: 28007982
 47. De Corte BJ, Matell MS. Temporal averaging across multiple response options: insight into the mechanisms underlying integration. *Animal cognition*. 2016; 19(2):329–342. <https://doi.org/10.1007/s10071-015-0935-4> PMID: 26520647
 48. Buhusi CV, Meck WH. What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*. 2005; 6(10):755–765. <https://doi.org/10.1038/nrn1764> PMID: 16163383
 49. Ludvig E, Sutton RS, Kehoe EJ, et al. Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. 2008.
 50. Gershman SJ, Moustafa AA, Ludvig EA. Time representation in reinforcement learning models of the basal ganglia. *Frontiers in computational neuroscience*. 2014; 7:194. <https://doi.org/10.3389/fncom.2013.00194> PMID: 24409138
 51. Petter EA, Gershman SJ, Meck WH. Integrating models of interval timing and reinforcement learning. *Trends in Cognitive Sciences*. 2018; 22(10):911–922. <https://doi.org/10.1016/j.tics.2018.08.004> PMID: 30266150
 52. Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, et al. Dopamine, affordance and active inference. *PLoS Computational Biology*. 2012; 8(1):e1002327. <https://doi.org/10.1371/journal.pcbi.1002327> PMID: 22241972
 53. Rammsayer TH. On dopaminergic modulation of temporal information processing. *Biological psychology*. 1993; 36(3):209–222. [https://doi.org/10.1016/0301-0511\(93\)90018-4](https://doi.org/10.1016/0301-0511(93)90018-4) PMID: 8260566

54. Ward RD, Kellendonk C, Simpson EH, Lipatova O, Drew MR, Fairhurst S, et al. Impaired timing precision produced by striatal D2 receptor overexpression is mediated by cognitive and motivational deficits. *Behavioral neuroscience*. 2009; 123(4):720. <https://doi.org/10.1037/a0016503> PMID: 19634929
55. Wiener M, Lohoff FW, Coslett HB. Double dissociation of dopamine genes and timing in humans. *Journal of cognitive neuroscience*. 2011; 23(10):2811–2821. <https://doi.org/10.1162/jocn.2011.21626> PMID: 21261454
56. Narayanan NS, Land BB, Solder JE, Deisseroth K, DiLeone RJ. Prefrontal D1 dopamine signaling is required for temporal control. *Proceedings of the National Academy of Sciences*. 2012; 109(50):20726–20731. <https://doi.org/10.1073/pnas.1211258109> PMID: 23185016
57. Parker KL, Chen KH, Kingyon JR, Cavanagh JF, Narayanan NS. D1-dependent 4 Hz oscillations and ramping activity in rodent medial frontal cortex during interval timing. *Journal of Neuroscience*. 2014; 34(50):16774–16783. <https://doi.org/10.1523/JNEUROSCI.2772-14.2014> PMID: 25505330
58. Espay AJ, Giuffrida JP, Chen R, Payne M, Mazzella F, Dunn E, et al. Differential response of speed, amplitude, and rhythm to dopaminergic medications in Parkinson's disease. *Movement Disorders*. 2011; 26(14):2504–2508. <https://doi.org/10.1002/mds.23893> PMID: 21953789
59. Popescu AT, Zhou MR, Poo Mm. Phasic dopamine release in the medial prefrontal cortex enhances stimulus discrimination. *Proceedings of the National Academy of Sciences*. 2016; 113(22):E3169–E3176. <https://doi.org/10.1073/pnas.1606098113> PMID: 27185946
60. Lou HC, Skewes JC, Thomsen KR, Overgaard M, Lau HC, Mouridsen K, et al. Dopaminergic stimulation enhances confidence and accuracy in seeing rapidly presented words. *Journal of vision*. 2011; 11(2):15–15. <https://doi.org/10.1167/11.2.15> PMID: 21346001
61. Andreou C, Moritz S, Veith K, Veckenstedt R, Naber D. Dopaminergic modulation of probabilistic reasoning and overconfidence in errors: a double-blind study. *Schizophrenia bulletin*. 2013; 40(3):558–565. <https://doi.org/10.1093/schbul/sbt064> PMID: 23661634
62. Andreou C, Bozikas VP, Luedtke T, Moritz S. Associations between visual perception accuracy and confidence in a dopaminergic manipulation study. *Frontiers in psychology*. 2015; 6:414. <https://doi.org/10.3389/fpsyg.2015.00414> PMID: 25932015
63. Sims CA. Implications of rational inattention. *Journal of Monetary Economics*. 2003; 50(3):665–690. [https://doi.org/10.1016/S0304-3932\(03\)00029-1](https://doi.org/10.1016/S0304-3932(03)00029-1)
64. Caplin A, Dean M. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*. 2015; 105(7):2183–2203. <https://doi.org/10.1257/aer.20140117>
65. Matějka F, McKay A. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*. 2015; 105(1):272–98. <https://doi.org/10.1257/aer.20130047>
66. Jackson DM, Andén NE, Dahlström A. A functional effect of dopamine in the nucleus accumbens and in some other dopamine-rich parts of the rat brain. *Psychopharmacologia*. 1975; 45(2):139–149. <https://doi.org/10.1007/BF00429052> PMID: 1215445
67. Carr GD, White NM. Effects of systemic and intracranial amphetamine injections on behavior in the open field: a detailed analysis. *Pharmacology Biochemistry and Behavior*. 1987; 27(1):113–122. [https://doi.org/10.1016/0091-3057\(87\)90485-0](https://doi.org/10.1016/0091-3057(87)90485-0) PMID: 3615534
68. Cousins MS, Salamone JD. Nucleus accumbens dopamine depletions in rats affect relative response allocation in a novel cost/benefit procedure. *Pharmacology Biochemistry and Behavior*. 1994; 49(1):85–91. [https://doi.org/10.1016/0091-3057\(94\)90460-X](https://doi.org/10.1016/0091-3057(94)90460-X)
69. Sokolowski J, Salamone J. The role of accumbens dopamine in lever pressing and response allocation: effects of 6-OHDA injected into core and dorsomedial shell. *Pharmacology Biochemistry and Behavior*. 1998; 59(3):557–566. [https://doi.org/10.1016/S0091-3057\(97\)00544-3](https://doi.org/10.1016/S0091-3057(97)00544-3) PMID: 9512057
70. Ikemoto S, Panksepp J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*. 1999; 31(1):6–41. [https://doi.org/10.1016/S0165-0173\(99\)00023-5](https://doi.org/10.1016/S0165-0173(99)00023-5) PMID: 10611493
71. Aberman J, Salamone JD. Nucleus accumbens dopamine depletions make rats more sensitive to high ratio requirements but do not impair primary food reinforcement. *Neuroscience*. 1999; 92(2):545–552. [https://doi.org/10.1016/S0306-4522\(99\)00004-4](https://doi.org/10.1016/S0306-4522(99)00004-4) PMID: 10408603
72. Salamone J, Wisniecki A, Carlson B, Correa M. Nucleus accumbens dopamine depletions make animals highly sensitive to high fixed ratio requirements but do not impair primary food reinforcement. *Neuroscience*. 2001; 105(4):863–870. [https://doi.org/10.1016/S0306-4522\(01\)00249-4](https://doi.org/10.1016/S0306-4522(01)00249-4) PMID: 11530224
73. Correa M, Carlson BB, Wisniecki A, Salamone JD. Nucleus accumbens dopamine and work requirements on interval schedules. *Behavioural brain research*. 2002; 137(1-2):179–187. [https://doi.org/10.1016/S0166-4328\(02\)00292-9](https://doi.org/10.1016/S0166-4328(02)00292-9) PMID: 12445723

74. Mingote S, Weber SM, Ishiwari K, Correa M, Salamone JD. Ratio and time requirements on operant schedules: effort-related effects of nucleus accumbens dopamine depletions. *European Journal of Neuroscience*. 2005; 21(6):1749–1757. <https://doi.org/10.1111/j.1460-9568.2005.03972.x> PMID: 15845103
75. Salamone JD, Correa M, Mingote SM, Weber SM. Beyond the reward hypothesis: alternative functions of nucleus accumbens dopamine. *Current opinion in pharmacology*. 2005; 5(1):34–41. <https://doi.org/10.1016/j.coph.2004.09.004> PMID: 15661623
76. Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*. 2007; 191(3):391–431. <https://doi.org/10.1007/s00213-006-0578-x> PMID: 17072591
77. Smith KS, Berridge KC, Aldridge JW. Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proceedings of the National Academy of Sciences*. 2011; 108(27):E255–E264. <https://doi.org/10.1073/pnas.1101920108> PMID: 21670308
78. Wassum KM, Ostlund SB, Balleine BW, Maidment NT. Differential dependence of Pavlovian incentive motivation and instrumental incentive learning processes on dopamine signaling. *Learning & memory*. 2011; 18(7):475–483. <https://doi.org/10.1101/lm.2229311> PMID: 21693635
79. Berridge KC. From prediction error to incentive salience: mesolimbic computation of reward motivation. *European Journal of Neuroscience*. 2012; 35(7):1124–1143. <https://doi.org/10.1111/j.1460-9568.2012.07990.x> PMID: 22487042
80. Beierholm U, Guitart-Masip M, Economides M, Chowdhury R, Düzal E, Dolan R, et al. Dopamine modulates reward-related vigor. *Neuropsychopharmacology*. 2013; 38(8):1495–1503. <https://doi.org/10.1038/npp.2013.48> PMID: 23419875
81. Qian N, Lipkin RM. A learning-style theory for understanding autistic behaviors. *Frontiers in Human Neuroscience*. 2011; 5:77. <https://doi.org/10.3389/fnhum.2011.00077> PMID: 21886617
82. Pellicano E, Burr D. When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in cognitive sciences*. 2012; 16(10):504–510. <https://doi.org/10.1016/j.tics.2012.08.009> PMID: 22959875
83. Sinha P, Kjelgaard MM, Gandhi TK, Tsourides K, Cardinaux AL, Pantazis D, et al. Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*. 2014; 111(42):15220–15225. <https://doi.org/10.1073/pnas.1416797111>
84. Palmer CJ, Lawson RP, Hohwy J. Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin*. 2017; 143(5):521. <https://doi.org/10.1037/bul0000097> PMID: 28333493
85. Adams RA, Perrinet LU, Friston K. Smooth pursuit and visual occlusion: active inference and oculomotor control in schizophrenia. *PloS One*. 2012; 7(10):e47502. <https://doi.org/10.1371/journal.pone.0047502> PMID: 23110076
86. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. *Frontiers in Psychiatry*. 2013; 4:47. <https://doi.org/10.3389/fpsy.2013.00047> PMID: 23750138
87. Corlett PR, Honey GD, Krystal JH, Fletcher PC. Glutamatergic model psychoses: prediction error, learning, and inference. *Neuropsychopharmacology*. 2011; 36(1):294. <https://doi.org/10.1038/npp.2010.163> PMID: 20861831
88. Marino RA, Levy R. Differential Effects of D1 and D2 Dopamine Agonists on Memory, Motivation, Learning and Response Time in Non-human Primates. *European Journal of Neuroscience*. 2018. PMID: 30326151
89. Posner MI. Attention in cognitive neuroscience: an overview. 1995.
90. Faisal AA, Selen LP, Wolpert DM. Noise in the nervous system. *Nature Reviews Neuroscience*. 2008; 9(4):292–303. <https://doi.org/10.1038/nrn2258> PMID: 18319728
91. Yuille AL, Bühlhoff HH. Bayesian decision theory and psychophysics. 1993.
92. Körding KP, Wolpert DM. Bayesian integration in sensorimotor learning. *Nature*. 2004; 427(6971):244. <https://doi.org/10.1038/nature02169> PMID: 14724638
93. Körding KP, Wolpert DM. Bayesian decision theory in sensorimotor control. *Trends in cognitive sciences*. 2006; 10(7):319–326. <https://doi.org/10.1016/j.tics.2006.05.003> PMID: 16807063
94. Doya K, Ishii S, Pouget A, Rao RP. Bayesian brain: Probabilistic approaches to neural coding. MIT press; 2007.
95. Wolpert DM. Probabilistic models in human sensorimotor control. *Human movement science*. 2007; 26(4):511–524. <https://doi.org/10.1016/j.humov.2007.05.005> PMID: 17628731
96. Körding KP, Wolpert DM. The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences*. 2004; 101(26):9839–9842. <https://doi.org/10.1073/pnas.0308394101> PMID: 15210973

97. Sun JZ, Wang GI, Goyal VK, Varshney LR. A framework for Bayesian optimality of psychophysical laws. *Journal of Mathematical Psychology*. 2012; 56(6):495–501. <https://doi.org/10.1016/j.jmp.2012.08.002>
98. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*. 2004; 27(12):712–719. <https://doi.org/10.1016/j.tins.2004.10.007> PMID: 15541511
99. Shannon CE. A mathematical theory of communication. *Bell system technical journal*. 1948; 27(3):379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
100. Cover TM, Thomas JA. *Elements of information theory*. John Wiley & Sons; 2012.
101. Mackowiak BA, Matejka F, Wiederholt M, et al. Survey: Rational Inattention, a Disciplined Behavioral Model. *CEPR Discussion Papers*; 2018.
102. Kurzban R, Duckworth A, Kable JW, Myers J. An opportunity cost model of subjective effort and task performance. *Behavioral and brain sciences*. 2013; 36(6):661–679. <https://doi.org/10.1017/S0140525X12003196> PMID: 24304775
103. Sims CR. Rate–distortion theory and human perception. *Cognition*. 2016; 152:181–198. <https://doi.org/10.1016/j.cognition.2016.03.020> PMID: 27107330
104. Averbeck BB, Costa VD. Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience*. 2017; 20(4):505. <https://doi.org/10.1038/nn.4506> PMID: 28352111
105. Shepard RN. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*. 1958; 55(6):509. <https://doi.org/10.1037/h0042354> PMID: 13563763
106. Luce RD. *Individual Choice Behavior: a Theoretical Analysis*. John Wiley and sons; 1959.
107. Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW. Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. *Journal of Neuroscience*. 2009; 29(48):15104–15114. <https://doi.org/10.1523/JNEUROSCI.3524-09.2009> PMID: 19955362
108. Schlagenhauf F, Rapp MA, Huys QJ, Beck A, Wüstenberg T, Deserno L, et al. Ventral striatal prediction error signaling is associated with dopamine synthesis capacity and fluid intelligence. *Human brain mapping*. 2013; 34(6):1490–1499. <https://doi.org/10.1002/hbm.22000> PMID: 22344813
109. Ward RD, Odum AL. Effects of prefeeding, intercomponent-interval food, and extinction on temporal discrimination and pacemaker rate. *Behavioural Processes*. 2006; 71(2-3):297–306. <https://doi.org/10.1016/j.beproc.2005.11.016> PMID: 16406376
110. Ward RD, Odum AL. Disruption of temporal discrimination and the choose-short effect. *Animal Learning & Behavior*. 2007; 35(1):60–70. <https://doi.org/10.3758/BF03196075> PMID: 17557392
111. Galtress T, Marshall AT, Kirkpatrick K. Motivation and timing: clues for modeling the reward system. *Behavioural processes*. 2012; 90(1):142–153. <https://doi.org/10.1016/j.beproc.2012.02.014> PMID: 22421220
112. Buhusi CV, Meck WH. Differential effects of methamphetamine and haloperidol on the control of an internal clock. *Behavioral neuroscience*. 2002; 116(2):291. <https://doi.org/10.1037/0735-7044.116.2.291> PMID: 11996314
113. Drew MR, Fairhurst S, Malapani C, Horvitz JC, Balsam PD. Effects of dopamine antagonists on the timing of two intervals. *Pharmacology Biochemistry and Behavior*. 2003; 75(1):9–15. [https://doi.org/10.1016/S0091-3057\(03\)00036-4](https://doi.org/10.1016/S0091-3057(03)00036-4) PMID: 12759108
114. Abner RT, Edwards T, Douglas A, Brunner D. Pharmacology of temporal cognition in two mouse strains. *International Journal of Comparative Psychology*. 2001; 14(3).
115. Çevik MÖ. Effects of methamphetamine on duration discrimination. *Behavioral Neuroscience*. 2003; 117(4):774. <https://doi.org/10.1037/0735-7044.117.4.774> PMID: 12931962
116. Matell MS, Bateson M, Meck WH. Single-trials analyses demonstrate that increases in clock speed contribute to the methamphetamine-induced horizontal shifts in peak-interval timing functions. *Psychopharmacology*. 2006; 188(2):201–212. <https://doi.org/10.1007/s00213-006-0489-x> PMID: 16937099
117. Killeen PR, Fetterman JG. A behavioral theory of timing. *Psychological review*. 1988; 95(2):274. <https://doi.org/10.1037/0033-295X.95.2.274> PMID: 3375401
118. Fetterman JG, Killeen PR. Adjusting the pacemaker. *Learning and Motivation*. 1991; 22(1):226–252. [https://doi.org/10.1016/0023-9690\(91\)90024-3](https://doi.org/10.1016/0023-9690(91)90024-3)
119. MacEwen D, Killeen P. The effects of rate and amount of reinforcement on the speed of the pacemaker in pigeons' timing behavior. *Learning & behavior*. 1991; 19(2):164–170. <https://doi.org/10.3758/BF03197872>

120. Morgan L, Killeen PR, Fetterman JG. Changing rates of reinforcement perturbs the flow of time. *Behavioural Processes*. 1993; 30(3):259–271. [https://doi.org/10.1016/0376-6357\(93\)90138-H](https://doi.org/10.1016/0376-6357(93)90138-H) PMID: [24896950](https://pubmed.ncbi.nlm.nih.gov/24896950/)
121. Killeen PR, Hall S, Bizo LA. A clock not wound runs down. *Behavioural Processes*. 1999; 45(1):129–139. [https://doi.org/10.1016/S0376-6357\(99\)00014-5](https://doi.org/10.1016/S0376-6357(99)00014-5) PMID: [24897532](https://pubmed.ncbi.nlm.nih.gov/24897532/)
122. Galtress T, Kirkpatrick K. Reward value effects on timing in the peak procedure. *Learning and Motivation*. 2009; 40(2):109–131. <https://doi.org/10.1016/j.lmot.2008.05.004>
123. Meck WH. Internal clock and reward pathways share physiologically similar information-processing stages. *Quantitative analyses of behavior: Biological determinants of reinforcement*. 2014; 7:121–38.
124. Meck WH, Church RM. Nutrients that modify the speed of internal clock and memory storage processes. *Behavioral Neuroscience*. 1987; 101(4):465. <https://doi.org/10.1037/0735-7044.101.4.465> PMID: [3651228](https://pubmed.ncbi.nlm.nih.gov/3651228/)
125. Roberts S. Isolation of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*. 1981; 7(3):242. PMID: [7252428](https://pubmed.ncbi.nlm.nih.gov/7252428/)
126. Mello GB, Soares S, Paton JJ. A scalable population code for time in the striatum. *Current Biology*. 2015; 25(9):1113–1122. <https://doi.org/10.1016/j.cub.2015.02.036> PMID: [25913405](https://pubmed.ncbi.nlm.nih.gov/25913405/)
127. Wang J, Narain D, Hosseini EA, Jazayeri M. Flexible timing by temporal scaling of cortical responses. *Nature neuroscience*. 2018; 21(1):102. <https://doi.org/10.1038/s41593-017-0028-6> PMID: [29203897](https://pubmed.ncbi.nlm.nih.gov/29203897/)
128. Gibbon J. Scalar expectancy theory and Weber's law in animal timing. *Psychological review*. 1977; 84(3):279. <https://doi.org/10.1037/0033-295X.84.3.279>
129. Church RM, Meck W. A concise introduction to scalar timing theory. *Functional and neural mechanisms of interval timing*. 2003; p. 3–22.
130. Staddon J. Some properties of spaced responding in pigeons. *Journal of the Experimental Analysis of Behavior*. 1965; 8(1):19–28. <https://doi.org/10.1901/jeab.1965.8-19> PMID: [14266504](https://pubmed.ncbi.nlm.nih.gov/14266504/)
131. Howard MW, Shankar KH. Neural scaling laws for an uncertain world. *Psychological review*. 2018; 125(1):47. <https://doi.org/10.1037/rev0000081> PMID: [29035080](https://pubmed.ncbi.nlm.nih.gov/29035080/)
132. Pardo-Vazquez JL, Castiñeiras-de Saa JR, Valente M, Damião I, Costa T, Vicente MI, et al. The mechanistic foundation of Weber's law. *Nature neuroscience*. 2019; p. 1–10. PMID: [31406366](https://pubmed.ncbi.nlm.nih.gov/31406366/)
133. Bizo LA, White KG. The behavioral theory of timing: Reinforcer rate determines pacemaker rate. *Journal of the Experimental Analysis of Behavior*. 1994; 61(1):19–33. <https://doi.org/10.1901/jeab.1994.61-19> PMID: [16812723](https://pubmed.ncbi.nlm.nih.gov/16812723/)
134. Bizo LA, White KG. Reinforcement context and pacemaker rate in the behavioral theory of timing. *Learning & behavior*. 1995; 23(4):376–382. <https://doi.org/10.3758/BF03198937>
135. Cabib S, Puglisi-Allegra S. Opposite responses of mesolimbic dopamine system to controllable and uncontrollable aversive experiences. *Journal of Neuroscience*. 1994; 14(5):3333–3340. <https://doi.org/10.1523/JNEUROSCI.14-05-03333.1994> PMID: [8182476](https://pubmed.ncbi.nlm.nih.gov/8182476/)
136. Cabib S, Puglisi-Allegra S. The mesoaccumbens dopamine in coping with stress. *Neuroscience & Biobehavioral Reviews*. 2012; 36(1):79–89. <https://doi.org/10.1016/j.neubiorev.2011.04.012> PMID: [21565217](https://pubmed.ncbi.nlm.nih.gov/21565217/)
137. Kacelnik A. Normative and descriptive models of decision making: time discounting and risk sensitivity. In: *CIBA foundation symposium*. Wiley Online Library; 1997. p. 51–70.
138. Daw ND, Touretzky DS. Behavioral considerations suggest an average reward TD model of the dopamine system. *Neurocomputing*. 2000; 32:679–684. [https://doi.org/10.1016/S0925-2312\(00\)00232-0](https://doi.org/10.1016/S0925-2312(00)00232-0)
139. Blanchard TC, Pearson JM, Hayden BY. Postreward delays and systematic biases in measures of animal temporal discounting. *Proceedings of the National Academy of Sciences*. 2013; 110(38):15491–15496. <https://doi.org/10.1073/pnas.1310446110>
140. Ainslie G. Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological bulletin*. 1975; 82(4):463. <https://doi.org/10.1037/h0076860> PMID: [1099599](https://pubmed.ncbi.nlm.nih.gov/1099599/)
141. Mikhael JG, Gershman SJ. Adapting the Flow of Time with Dopamine. *Journal of neurophysiology*. 2019. <https://doi.org/10.1152/jn.00817.2018> PMID: [30864882](https://pubmed.ncbi.nlm.nih.gov/30864882/)
142. Manohar SG, Chong TTJ, Apps MA, Batla A, Stamelou M, Jarman PR, et al. Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology*. 2015; 25(13):1707–1716. <https://doi.org/10.1016/j.cub.2015.05.038> PMID: [26096975](https://pubmed.ncbi.nlm.nih.gov/26096975/)
143. Manohar SG, Muhammed K, Fallon SJ, Husain M. Motivation dynamically increases noise resistance by internal feedback during movement. *Neuropsychologia*. 2019; 123:19–29. <https://doi.org/10.1016/j.neuropsychologia.2018.07.011> PMID: [30005926](https://pubmed.ncbi.nlm.nih.gov/30005926/)
144. Otto AR, Daw ND. The opportunity cost of time modulates cognitive effort. *Neuropsychologia*. 2019; 123:92–105. <https://doi.org/10.1016/j.neuropsychologia.2018.05.006> PMID: [29750987](https://pubmed.ncbi.nlm.nih.gov/29750987/)

145. Zahrt J, Taylor JR, Mathew RG, Arnsten AF. Supranormal stimulation of D1 dopamine receptors in the rodent prefrontal cortex impairs spatial working memory performance. *Journal of neuroscience*. 1997; 17(21):8528–8535. <https://doi.org/10.1523/JNEUROSCI.17-21-08528.1997> PMID: 9334425
146. Vijayraghavan S, Wang M, Birnbaum SG, Williams GV, Arnsten AF. Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nature neuroscience*. 2007; 10(3):376. <https://doi.org/10.1038/nn1846>
147. Cools R, D'Esposito M. Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biological psychiatry*. 2011; 69(12):e113–e125. <https://doi.org/10.1016/j.biopsych.2011.03.028> PMID: 21531388
148. Coull JT, Hwang HJ, Leyton M, Dagher A. Dopaminergic modulation of motor timing in healthy volunteers differs as a function of baseline DA precursor availability. *Timing & Time Perception*. 2013; 1(1):77–98. <https://doi.org/10.1163/22134468-00002005>
149. Meder D, Herz DM, Rowe JB, Lehericy S, Siebner HR. The role of dopamine in the brain—lessons learned from Parkinson's disease. *NeuroImage*. 2018. PMID: 30465864
150. 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*. 2009; 4(8):e6507. <https://doi.org/10.1371/journal.pone.0006507> PMID: 19654866
151. Soares-Cunha C, Coimbra B, David-Pereira A, Borges S, Pinto L, Costa P, et al. Activation of D2 dopamine receptor-expressing neurons in the nucleus accumbens increases motivation. *Nature communications*. 2016; 7:11829. <https://doi.org/10.1038/ncomms11829> PMID: 27337658
152. Hamilos AE, Spedicato G, Hong Y, Sun F, Li Y, Assad JA. Dynamic dopaminergic activity controls the timing of self-timed movement. *bioRxiv*. 2020.
153. Hamilos AE, Assad JA. Application of a unifying reward-prediction error (RPE)-based framework to explain underlying dynamic dopaminergic activity in timing tasks. *bioRxiv*. 2020.
154. Howe MW, Tierney PL, Sandberg SG, Phillips PE, Graybiel AM. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature*. 2013; 500(7464):575. <https://doi.org/10.1038/nature12475> PMID: 23913271
155. Kobayashi S, Schultz W. Influence of reward delays on responses of dopamine neurons. *Journal of neuroscience*. 2008; 28(31):7837–7846. <https://doi.org/10.1523/JNEUROSCI.1600-08.2008> PMID: 18667616
156. Stuber GD, Klanker M, de Ridder B, Bowers MS, Joosten RN, Feenstra MG, et al. Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science*. 2008; 321(5896):1690–1692. <https://doi.org/10.1126/science.1160873> PMID: 18802002
157. Fligel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, et al. A selective role for dopamine in stimulus–reward learning. *Nature*. 2011; 469(7328):53. <https://doi.org/10.1038/nature09588> PMID: 21150898
158. Lloyd K, Dayan P. Tamping ramping: Algorithmic, implementational, and computational explanations of phasic dopamine signals in the accumbens. *PLoS computational biology*. 2015; 11(12):e1004622. <https://doi.org/10.1371/journal.pcbi.1004622> PMID: 26699940
159. Gershman SJ. Dopamine ramps are a consequence of reward prediction errors. *Neural computation*. 2014; 26(3):467–471. https://doi.org/10.1162/NECO_a_00559 PMID: 24320851
160. Kim HR, Malik AN, Mikhael JG, Bech P, Tsutsui-Kimura I, Sun F, et al. A unified framework for dopamine signals across timescales. *Cell*. 2020. <https://doi.org/10.1016/j.cell.2020.11.013> PMID: 33248024
161. Mikhael JG, Kim HR, Uchida N, Gershman SJ. Ramping and State Uncertainty in the Dopamine Signal. *bioRxiv*. 2019; p. 805366.
162. Hamid AA, Frank MJ, Moore CI. Dopamine waves as a mechanism for spatiotemporal credit assignment. *bioRxiv*. 2019; p. 729640.
163. Brozoski TJ, Brown RM, Rosvold H, Goldman PS. Cognitive deficit caused by regional depletion of dopamine in prefrontal cortex of rhesus monkey. *Science*. 1979; 205(4409):929–932. <https://doi.org/10.1126/science.112679> PMID: 112679
164. Daniel D, Weinberger D, Jones D, Zigun J, Coppola R, Handel S, et al. The effect of amphetamine on regional cerebral blood flow during cognitive activation in schizophrenia. *Journal of Neuroscience*. 1991; 11(7):1907–1917. <https://doi.org/10.1523/JNEUROSCI.11-07-01907.1991> PMID: 2066768
165. Lange KW, Robbins T, Marsden C, James M, Owen A, Paul G. L-dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. *Psychopharmacology*. 1992; 107(2-3):394–404. <https://doi.org/10.1007/BF02245167> PMID: 1615139
166. Williams GV, Goldman-Rakic PS. Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature*. 1995; 376(6541):572. <https://doi.org/10.1038/376572a0> PMID: 7637804

167. Dodson PD, Dreyer JK, Jennings KA, Syed ECJ, Wade-Martins R, Cragg SJ, et al. Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proceedings of the National Academy of Sciences*. 2016; p. 201515941. <https://doi.org/10.1073/pnas.1515941113> PMID: 27001837
168. Howe MW, Dombeck DA. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature*. 2016; 535(7613):505. <https://doi.org/10.1038/nature18942> PMID: 27398617
169. Harrison AA, Everitt BJ, Robbins TW. Central 5-HT depletion enhances impulsive responding without affecting the accuracy of attentional performance: interactions with dopaminergic mechanisms. *Psychopharmacology*. 1997; 133(4):329–342. <https://doi.org/10.1007/s002130050410> PMID: 9372531
170. Winstanley CA, Theobald DE, Dalley JW, Cardinal RN, Robbins TW. Double dissociation between serotonergic and dopaminergic modulation of medial prefrontal and orbitofrontal cortex during a test of impulsive choice. *Cerebral Cortex*. 2006; 16(1):106–114. <https://doi.org/10.1093/cercor/bhi088> PMID: 15829733
171. Winstanley CA, Theobald DE, Dalley JW, Robbins TW. Interactions between serotonin and dopamine in the control of impulsive choice in rats: therapeutic implications for impulse control disorders. *Neuropsychopharmacology*. 2005; 30(4):669–682. <https://doi.org/10.1038/sj.npp.1300610> PMID: 15688093
172. Pine A, Shiner T, Seymour B, Dolan RJ. Dopamine, time, and impulsivity in humans. *Journal of Neuroscience*. 2010; 30(26):8888–8896. <https://doi.org/10.1523/JNEUROSCI.6028-09.2010> PMID: 20592211
173. Kayser AS, Allen DC, Navarro-Cebrian A, Mitchell JM, Fields HL. Dopamine, corticostriatal connectivity, and intertemporal choice. *Journal of Neuroscience*. 2012; 32(27):9402–9409. <https://doi.org/10.1523/JNEUROSCI.1180-12.2012> PMID: 22764248
174. Castrellon JJ, Meade J, Greenwald L, Hurst K, Samanez-Larkin GR. Dopaminergic modulation of reward discounting: a systematic review and meta-analysis. *bioRxiv*. 2020. PMID: 33215269
175. Boot N, Baas M, van Gaal S, Cools R, De Dreu CK. Creative cognition and dopaminergic modulation of fronto-striatal networks: Integrative review and research agenda. *Neuroscience & Biobehavioral Reviews*. 2017; 78:13–23. <https://doi.org/10.1016/j.neubiorev.2017.04.007> PMID: 28419830
176. Faust-Socher A, Kenett YN, Cohen OS, Hassin-Baer S, Inzelberg R. Enhanced creative thinking under dopaminergic therapy in Parkinson disease. *Annals of neurology*. 2014; 75(6):935–942. <https://doi.org/10.1002/ana.24181> PMID: 24816898
177. Graybiel AM. The basal ganglia. *Current Biology*. 2000; 10(14):R509–R511. [https://doi.org/10.1016/S0960-9822\(00\)00593-5](https://doi.org/10.1016/S0960-9822(00)00593-5) PMID: 10899013
178. Albin RL, Young AB, Penney JB. The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*. 1989; 12:366–375. [https://doi.org/10.1016/0166-2236\(89\)90074-X](https://doi.org/10.1016/0166-2236(89)90074-X)
179. DeLong MR. Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences*. 1990; 13:281–285. [https://doi.org/10.1016/0166-2236\(90\)90110-V](https://doi.org/10.1016/0166-2236(90)90110-V) PMID: 1695404
180. Shen W, Flajolet M, Greengard P, Surmeier DJ. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science*. 2008; 321(5890):848–851. <https://doi.org/10.1126/science.1160575>
181. Dreyer JK, Herrik KF, Berg RW, Hounsgaard JD. Influence of phasic and tonic dopamine release on receptor activation. *The Journal of Neuroscience*. 2010; 30(42):14273–14283. <https://doi.org/10.1523/JNEUROSCI.1894-10.2010> PMID: 20962248
182. Frank M. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Cognitive Neuroscience, Journal of*. 2005; 17(1):51–72. <https://doi.org/10.1162/0898929052880093> PMID: 15701239
183. Drew MR, Simpson EH, Kellendonk C, Herzberg WG, Lipatova O, Fairhurst S, et al. Transient overexpression of striatal D2 receptors impairs operant motivation and interval timing. *Journal of Neuroscience*. 2007; 27(29):7731–7739. <https://doi.org/10.1523/JNEUROSCI.1736-07.2007> PMID: 17634367
184. Coull JT, Cheng RK, Meck WH. Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology*. 2011; 36(1):3–25. <https://doi.org/10.1038/npp.2010.113> PMID: 20668434
185. De Corte BJ, Wagner LM, Matell MS, Narayanan NS. Striatal dopamine and the temporal control of behavior. *Behavioural brain research*. 2019; 356:375–379. <https://doi.org/10.1016/j.bbr.2018.08.030> PMID: 30213664
186. Marr D. *Vision: A computational investigation into the human representation and processing of visual information*, Henry Holt and Co. Inc, New York, NY. 1982;2(4.2).
187. Bjorklund A, Lindvall O. Dopamine-containing systems in the CNS; 1984.
188. Le Moal M, Simon H. Mesocorticolimbic dopaminergic network: functional and regulatory roles. *Physiological reviews*. 1991; 71(1):155–234. <https://doi.org/10.1152/physrev.1991.71.1.155> PMID: 1986388

189. Balci F, Wiener M, Cavdaroglu B, Coslett HB. Epistasis effects of dopamine genes on interval timing and reward magnitude in humans. *Neuropsychologia*. 2013; 51(2):293–308. <https://doi.org/10.1016/j.neuropsychologia.2012.08.002> PMID: 22903038
190. Gershman SJ, Tzovaras BG. Dopaminergic genes are associated with both directed and random exploration. *Neuropsychologia*. 2018; 120:97–104. <https://doi.org/10.1016/j.neuropsychologia.2018.10.009> PMID: 30347192
191. Oberauer K, Farrell S, Jarrold C, Lewandowsky S. What limits working memory capacity? *Psychological bulletin*. 2016; 142(7):758. <https://doi.org/10.1037/bul0000046> PMID: 26950009
192. Bossaerts P, Yadav N, Murawski C. Uncertainty and computational complexity. *Philosophical Transactions of the Royal Society B*. 2019; 374(1766):20180138. <https://doi.org/10.1098/rstb.2018.0138> PMID: 30966921
193. Musslick S, Dey B, Ozcimder K, Patwary MMA, Willke TL, Cohen JD. Parallel processing capability versus efficiency of representation in neural networks. *Network*. 2016;8(7).
194. Gailliot MT, Baumeister RF. The physiology of willpower: Linking blood glucose to self-control. *Personality and social psychology review*. 2007; 11(4):303–327. <https://doi.org/10.1177/1088868307303030> PMID: 18453466
195. Shenhav A, Musslick S, Lieder F, Kool W, Griffiths TL, Cohen JD, et al. Toward a rational and mechanistic account of mental effort. *Annual review of neuroscience*. 2017; 40:99–124. <https://doi.org/10.1146/annurev-neuro-072116-031526> PMID: 28375769
196. Behrens TE, Woolrich MW, Walton ME, Rushworth MF. Learning the value of information in an uncertain world. *Nature neuroscience*. 2007; 10(9):1214–1221. <https://doi.org/10.1038/nn1954> PMID: 17676057
197. Yu AJ, Dayan P. Uncertainty, neuromodulation, and attention. *Neuron*. 2005; 46(4):681–692. <https://doi.org/10.1016/j.neuron.2005.04.026> PMID: 15944135
198. Starkweather CK, Babayan BM, Uchida N, Gershman SJ. Dopamine reward prediction errors reflect hidden-state inference across time. *Nature Neuroscience*. 2017; 20(4):581–589. <https://doi.org/10.1038/nn.4520> PMID: 28263301
199. Starkweather CK, Gershman SJ, Uchida N. The Medial Prefrontal Cortex Shapes Dopamine Reward Prediction Errors under State Uncertainty. *Neuron*. 2018; 98:616–629. <https://doi.org/10.1016/j.neuron.2018.03.036> PMID: 29656872
200. Fiorillo CD, Tobler PN, Schultz W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*. 2003; 299(5614):1898–1902. <https://doi.org/10.1126/science.1077349> PMID: 12649484
201. Hart AS, Clark JJ, Phillips PE. Dynamic shaping of dopamine signals during probabilistic Pavlovian conditioning. *Neurobiology of Learning and Memory*. 2015; 117:84–92. <https://doi.org/10.1016/j.nlm.2014.07.010> PMID: 25172480
202. Humphries MD, Khamassi M, Gurney K. Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in neuroscience*. 2012; 6:9. <https://doi.org/10.3389/fnins.2012.00009> PMID: 22347155
203. Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD. Humans use directed and random exploration to solve the explore–exploit dilemma. *Journal of Experimental Psychology: General*. 2014; 143(6):2074. <https://doi.org/10.1037/a0038199> PMID: 25347535
204. Gershman SJ. Deconstructing the human algorithms for exploration. *Cognition*. 2018; 173:34–42. <https://doi.org/10.1016/j.cognition.2017.12.014> PMID: 29289795
205. Myerson J, Green L. Discounting of delayed rewards: Models of individual choice. *Journal of the experimental analysis of behavior*. 1995; 64(3):263–276. <https://doi.org/10.1901/jeab.1995.64-263> PMID: 16812772
206. Kacelnik A, Bateson M. Risky theories—the effects of variance on foraging decisions. *American Zoologist*. 1996; 36(4):402–434. <https://doi.org/10.1093/icb/36.4.402>
207. Bernoulli D. Exposition of a new theory on the measurement of risk. *Econometrica*. 1954; 22(1):23–36. <https://doi.org/10.2307/1909829>
208. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*. 1979; p. 263–291. <https://doi.org/10.2307/1914185>