Impulsivity as Bayesian inference under dopaminergic control

John G. Mikhael[1,2*] & Samuel J. Gershman[3,4]

[1]Program in Neuroscience, Harvard Medical School, Boston, MA 02115

[2]MD-PhD Program, Harvard Medical School, Boston, MA 02115

[3]Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA 02138

[4]Center for Brains, Minds and Machines, Massachusetts Institute of Technology, Cambridge, MA 02139

*Corresponding Author

Correspondence: john_mikhael@hms.harvard.edu

Abstract

Bayesian models successfully account for several of dopamine (DA)'s effects on contextual calibration in interval timing and reward estimation. In these models, DA controls the precision of stimulus encoding, which is weighed against contextual information when making decisions. When DA levels are high, the animal relies more heavily on the (highly precise) stimulus encoding, whereas when DA levels are low, the context affects decisions more strongly. Here, we extend this idea to intertemporal choice tasks, in which agents must choose between small rewards delivered soon and large rewards delivered later. Beginning with the principle that animals will seek to maximize their reward rates, we show that the Bayesian model predicts a number of curious empirical findings. First, the model predicts that higher DA levels should normally promote selection of the larger/later option, which is often taken to imply that DA decreases 'impulsivity.' However, if the temporal precision is sufficiently decreased, higher DA levels should have the opposite effect—promoting selection of the smaller/sooner option (more impulsivity). Second, in both cases, high enough levels of DA can result in preference reversals. Third, selectively decreasing the temporal precision, without manipulating DA, should promote selection of the larger/later option. Fourth, when a different post-reward delay is associated with each option, animals will not learn the option-delay contingencies, but this learning can be salvaged when the post-reward delays are made more salient. Finally, the Bayesian model predicts a correlation between behavioral phenotypes: Animals that are better timers will also appear less impulsive.

Keywords: dopamine, Bayesian inference, precision, impulsivity, interval timing, central tendency, post-reward delay

Significance Statement

Does dopamine make animals more or less impulsive? Though impulsivity features prominently in several dopamine-related conditions, how dopamine actually influences impulsivity has remained unclear. In intertemporal choice tasks (ITCs), wherein animals must choose between small rewards delivered soon and large rewards delivered later, administering dopamine makes animals more willing to wait for larger/later rewards in some conditions (consistent with lower impulsivity), but less willing in others. We hypothesize that dopamine does not necessarily influence impulsivity at all, but rather gates the influence of contextual information during decision making. We show that this account explains an array of curious findings in ITCs, including the seemingly conflicting results above. Our work encourages a reexamination of ITCs as a method for assessing impulsivity.

Introduction

The neuromodulator dopamine (DA) has been repeatedly associated with choice impulsivity, the tendency to prioritize short-term over long-term reward. Impulsive behaviors characterize a number of DA-related psychiatric conditions (1), such as attention-deficit/hyperactivity disorder (2–6), schizophrenia (7, 8), addiction (9, 10), and dopamine dysregulation syndrome (11, 12). Furthermore, direct pharmacological manipulation of DA in humans (13, 14) and rodents (15, 16) has corroborated a relationship between DA and impulsivity. The standard approach to measuring impulsive choice is the intertemporal choice task (ITC), in which subjects choose between a small reward delivered soon and a large reward delivered later (17). A subject's preference for the smaller/sooner option is often taken as a measure of their impulsivity, or the extent to which they discount future rewards (18–21).

In the majority of animal studies, higher DA levels have been found to promote selection of the larger/later option (inhibiting impulsivity). However, the inference that DA inhibits impulsivity has been challenged in recent years, in part because, when ITCs are administered to humans, DA seems to *promote* impulsivity (22). Perhaps relevant to this contrast is that, while impulsive choices in humans are assessed through hypothetical situations ('Would you prefer $1 now or $10 in one month?'), ITCs in animals more closely resemble reinforcement learning tasks involving many trials of experienced rewards and delays. Complicating this picture further, the effect of DA, even within animal studies, is not straightforward. While in most studies, DA appears to decrease impulsivity, DA has been found to systematically increase impulsivity under some conditions, such as when the delay period is uncued (16) or when different delays for the larger/later option are presented in decreasing order across training blocks (23).

Animal behavior in ITCs can be reinterpreted from a reinforcement learning perspective. With repeated trials of the same task, an optimal agent can learn to maximize its total accumulated rewards by estimating the reward rate for each option (reward magnitude divided by total trial duration) and choosing the option with the higher reward rate. Thus if the larger/later option has a sufficiently large reward or sufficiently short delay, it will be the optimal choice. However, if its reward were sufficiently small or its delay sufficiently long, the smaller/sooner option may be the superior choice instead, without any assumption of 'discounting.' Under this view, animals do not necessarily discount future rewards at all, but rather make choices based on a reward-rate computation. The notion of true impulsivity in ITCs has persisted, however, because animals tend to choose the smaller/sooner option even when it objectively yields fewer rewards over many trials.

To address the question of whether animals simply compare reward rates, a body of theoretical and ex-

perimental work has demonstrated that the suboptimal tendency to choose the smaller/sooner option is better explained by *temporal* biases than by biases of choice (24–26; see also 27). This work has shown that animals behave in a way consistent with maximizing their reward rates, but they underestimate the elapsed time—and in particular, the periods after receiving the reward and before beginning the next trial. Thus animals estimate the reward rates for each option based largely on the pre-reward delays. This bias disproportionately benefits the smaller/sooner option, which has a much shorter pre-reward delay. As a result, the animals make choices that can be interpreted as impulsive. Said differently, animals disproportionately underestimate the total trial duration for the smaller/sooner option compared to the larger/later option, making the former more appealing. While this discounting-free view derives animal behavior from a normative framework (maximizing reward rates), how and why DA modulates choice preferences remains the subject of much speculation.

In this paper, we build on recent theoretical work that cast DA in a Bayesian light (28, 29). Here, DA controls the precision with which cues are internally represented, which in turn controls the extent to which the animal's estimates of the cues are influenced by context. In Bayesian terms, which we discuss below, DA controls the precision of the likelihood relative to that of the prior (the context). This framework predicts a well-replicated result in the interval timing literature, referred to as the 'central tendency' effect: When temporal intervals of different lengths are reproduced under DA depletion (e.g., in unmedicated Parkinson's patients), shorter intervals tend to be overproduced and longer intervals tend to be underproduced, and DA repletion rescues accurate timing (30–32). We recently extended this framework to the representation of reward estimates (33). In this case, the Bayesian framework predicts that DA should tip the exploration-exploitation balance toward exploitation, in line with empirical findings (34–36, but see 37, 38).

We show here that, under the Bayesian theory, DA should promote behaviors consistent with lower impulsivity in the standard ITC task (selection of the larger/later option), but should have the opposite effect when the temporal precision of the delay period is selectively and sufficiently reduced. In both cases, high enough levels of DA should elicit preference reversals, and not only an amplification of the current preference. Furthermore, in manipulations of temporal precision, if animals are more likely to select the larger/later option at baseline, DA administration will tend to reverse that preference (promote the smaller/sooner option), and vice versa. We show that animals should not learn the contingencies between options and their post-reward delays, but that this learning can be salvaged if the post-reward delays are made more salient. Finally, we show that animals that display more precise behaviors in interval timing tasks should also appear less impulsive.

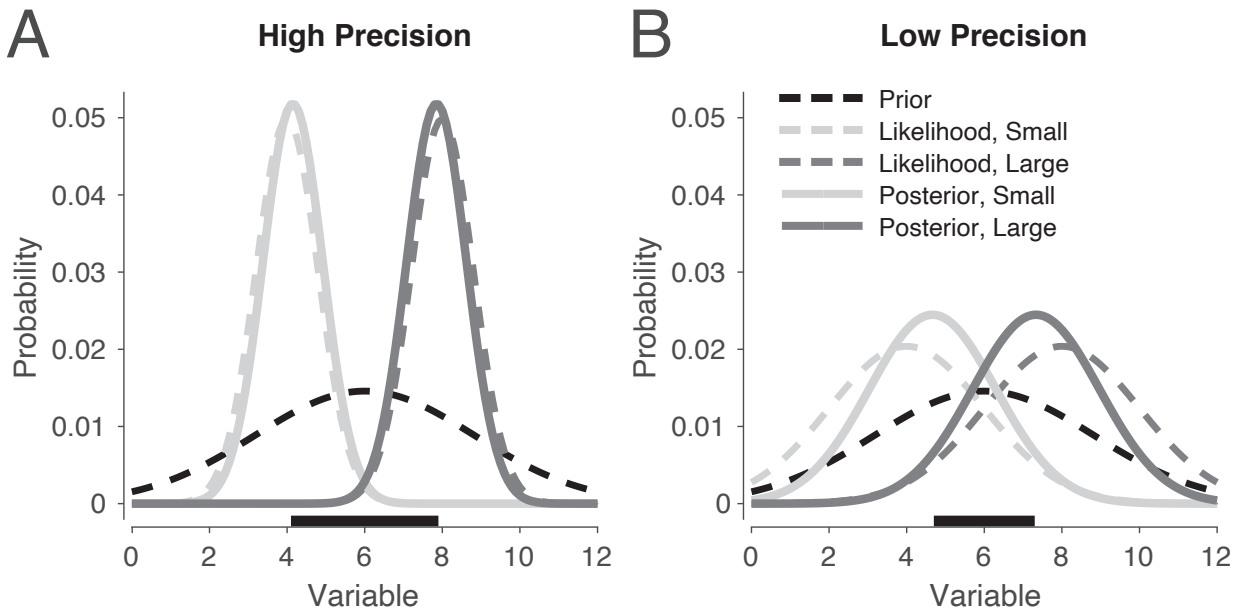

Figure 1: Contextual influence is stronger when the encoding precision is low. Plotted are the distributions for two signals, one small and the other large. (A) When the encoding precision is high compared to the prior precision, the posteriors do not deviate significantly from the likelihood. (B) As the encoding precision decreases, the posteriors migrate toward the prior. The horizontal black segments illustrate the difference in posterior means under high vs. low precision.

Results

The Bayesian theory of dopamine

An agent wishing to encode information about some cue must contend with noise at every level, including the information source (which is seldom deterministic), storage (synapses are noisy), and signaling (neurons are noisy; 39). We can formalize the noisy encoding as a mapping from an input signal (e.g., experienced reward) to a distribution over output signals (e.g., firing rates). For the purposes of this paper, we will remain agnostic about the specific neural implementation of the mapping, and instead discuss it in abstract terms. Thus a noisy encoding of some variable can be represented by a distribution over values: Tight distributions correspond to encodings with low noise (Fig. 1A), whereas wide distributions correspond to encodings with high noise (Fig. 1B).

Consider, then, a scenario in which an animal must estimate the average yield of a reward source from noisy samples. Because of the animal's uncertainty about the average yield (the encoding distribution has non-zero spread), its final estimate can be improved by utilizing other sources of information. For example, if the nearby reward sources tend to yield large rewards, then the animal should form an optimistic estimate of

the reward source's average yield. Similarly, if nearby reward sources yield small rewards, then the animal should form a pessimistic estimate. Formally, the contextual information can be used to construct a prior distribution over average yield, and the encoding distribution can be used to construct a likelihood function for evaluating the consistency between the encoded information and a hypothetical average yield. Bayes' rule stipulates that the animal's final probabilistic estimate should reflect the product of the likelihood and prior:

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

referred to as the posterior distribution. Here, μ is the variable being estimated (the reward yield), m is the stored value, $p(m|\mu)$ is the likelihood, and $p(\mu)$ is the prior. For simplicity, we take these distributions to be Gaussian throughout. Under standard assumptions for Gaussian distributions, the estimate $\hat{\mu}$ corresponds to the posterior mean:

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

Here, μ_0, λ_0, μ, and λ represent the prior mean, prior precision, likelihood mean, and encoding precision, respectively. In words, the agent takes a weighted average of the prior mean μ_0 and the likelihood mean μ—weighted by their respective precisions λ_0 and λ after normalization—to produce its estimate, the posterior mean $\hat{\mu}$. Intuitively, the tighter each distribution, the more it pulls the posterior mean in its direction.

The Bayesian theory of DA asserts that DA controls the encoding precision λ, where the prior here represents the distribution of stimuli (i.e., the context). Thus when DA is high, the estimate $\hat{\mu}$ does not heavily depend on contextual information, whereas when it is low, Bayesian migration of the estimate to the prior is strong (compare Fig. 1A and B). Shi et al. (32) have applied this theory to interval timing and shown that it predicts DA's effects on the central tendency: Parkinson's patients who are on their medication will have high λ, qualitatively corresponding to Fig. 1A. Then the temporal estimates for the short and long durations will be very close to their true values (here, 4 and 8 seconds). On the other hand, patients who are off their medication will have low λ, corresponding to Fig. 1B. Thus the estimates for both durations will migrate toward the prior mean, or the average of the two durations. In other words, the estimate for the short duration will be overproduced, and the estimate for the long duration will be underproduced.

The Bayesian model can also be applied to reward magnitudes. Imagine a bandit task in which an agent samples from two reward sources, one yielding small rewards on average and the other yielding large rewards on average. Under lower levels of DA, the central tendency should decrease the difference between the two reward estimates (compare lengths of black segments on the x-axis in Fig. 1A and B). Under standard models of action selection, animals are more likely to choose the large option when the difference between

the two estimates is large (a tendency to exploit the larger option; see Methods), and become more and more likely to sample other options as the difference decreases (a tendency to explore other options). This means that lower levels of DA should shift the exploration-exploitation trade-off toward exploration (selecting the smaller reward), as empirically observed (34–36, but see 37, 38). There is some behavioral evidence to suggest reward magnitude learning is indeed influenced by context in a way that follows our Bayesian framework (40, 41), although DA's role in this framework has not been examined directly for the domain of rewards. In addition, notice here that the Bayesian framework subsumes the 'gain control' theory of DA, in which high DA levels have been hypothesized to amplify the difference between reward estimates during decision making (42–45, see Methods).

Finally, we can compare the degree of the central tendency in temporal and reward estimation, which will be important in the next section. Empirically, the central tendency in temporal tasks is normally weak. While it can be unmasked in healthy subjects (46–50) and animals (51), it is most evident in unmedicated Parkinson's patients (30), in whom the DA deficiency is profound. This implies a significant asymmetry at baseline: While decreasing the DA levels will have a strong behavioral signature (the central tendency), the effect of increased DA levels will be small (due to a 'ceiling effect,' in which the central tendency will continue to be weak). On the other hand, both increases and decreases to the DA level substantially affect the exploration-exploitation trade-off (34–36, 52, 53). This suggests a more significant central tendency for rewards at baseline, which can be amplified or mitigated by DA manipulations. Below we will find that DA's effect in ITCs will depend on its relative contribution to each of the reward estimates and temporal estimates at baseline. Driven by the empirical observations, we take the baseline central tendency to be weaker in the domain of timing than in the domain of rewards.

Dopamine and intertemporal choice

ITCs involve choosing between a small reward delivered soon, and a large reward delivered later. In these tasks, the smaller/sooner delay is held fixed (and is often zero, resulting in immediate reward), while the larger/later delay is varied across blocks. When the delays are equal, animals will overwhelmingly choose the larger option, but as the delay for the larger option gets longer, animals become more and more likely to choose the smaller/sooner option (Fig. 3). This shift toward the smaller/sooner option has traditionally been explained in terms of reward discounting: The promise of a future reward is less valuable than that same reward delivered immediately, and becomes even less valuable as the delay increases. In other words, future rewards are discounted in proportion to the delay required to receive them. Previous computational models

have shown this reward discounting to be well-described by a hyperbolic (or quasi-hyperbolic) function (21, 54).

A competing line of thought is that animals seek to maximize their reward rates (or equivalently, the total accumulated rewards in the task; 24, 25, 27), but are limited by a significant underestimation of the post-reward delays in the task (26). On this view, animals compute the reward rate for each option—i.e., the undiscounted reward magnitude divided by the total trial time—but base the trial time largely on the pre-reward delay. This causes the reward rate for the smaller/sooner option to be disproportionately overestimated compared to that of the larger/later option. This view, much like the discounting view, predicts that animals will choose the larger/later option when its delay is short, but will gradually begin to prefer the smaller/sooner option as the delay is increased. Furthermore, the smaller/sooner option will be preferred in some cases even when it yields a lower reward rate, although this is due to a temporal bias (underestimation of post-reward delays), rather than a choice bias (reward discounting).

While the reward-rate interpretation can accommodate the aspects of the data explained by the discounting model (see Methods), it also captures aspects of animal behavior where the discounting model fails. In particular, Blanchard et al. (26) examined the effect of post-reward delays on behavior. Under the discounting model, behavior depends only on the reward magnitudes and *pre*-reward delays (over which the discounting occurs), and thus should be invariant to changes in the post-reward delays. The authors, however, found that monkeys modified their choices in line with a reward-rate computation, which must take into account both pre- and post-reward delays when computing the total trial time. Interestingly, the best fit to the data required that the post-reward delays be underestimated by about a factor of four, consistent with a bias of timing rather than a bias of choice in explaining animal behavior in ITCs. In what follows, we adopt the reward-rate interpretation in examining DA's role in ITCs.

Given DA's effects on reward estimates and durations, it is not surprising that DA would influence behavior in ITCs, where the agent's task is to maximize the ratio of these two, the reward rate $\bar{R}$:

$$\bar{R} = \frac{w_r \mu_r + (1 - w_r)\mu_{r0}}{w_t \mu_t + (1 - w_t)\mu_{t0}}, \tag{3}$$

which follows from Eq. (2). Here, $w_r = \frac{\lambda_r}{\lambda_r + \lambda_{r0}}$, and μ_{r0}, λ_{r0}, μ_r, and λ_r in the numerator represent the prior mean, prior precision, encoding distribution mean, and encoding distribution precision in the domain of rewards, respectively, and similarly for the domain of time in the denominator. The Bayesian framework captures the hyperbolic pattern observed under the discounting model (see Methods).

The ultimate effect of DA will depend on its relative contribution to the numerator and denominator: In the numerator, a stronger central tendency for the estimated rewards causes an overestimation of the smaller reward and an underestimation of the larger reward, thus promoting selection of the smaller/sooner option compared to baseline. Because DA masks the central tendency, its effect on the numerator is to promote selecting the larger/later option (Fig. 2, top arrow). On the other hand, in the denominator, a stronger central tendency for the estimated durations causes an overestimation of the sooner duration and an underestimation of the later duration, thus promoting selection of the larger/later option. Because DA masks the central tendency, its effect on the denominator is to promote selecting the smaller/sooner option—the opposite of its effect in the numerator (Fig. 2, bottom arrow).

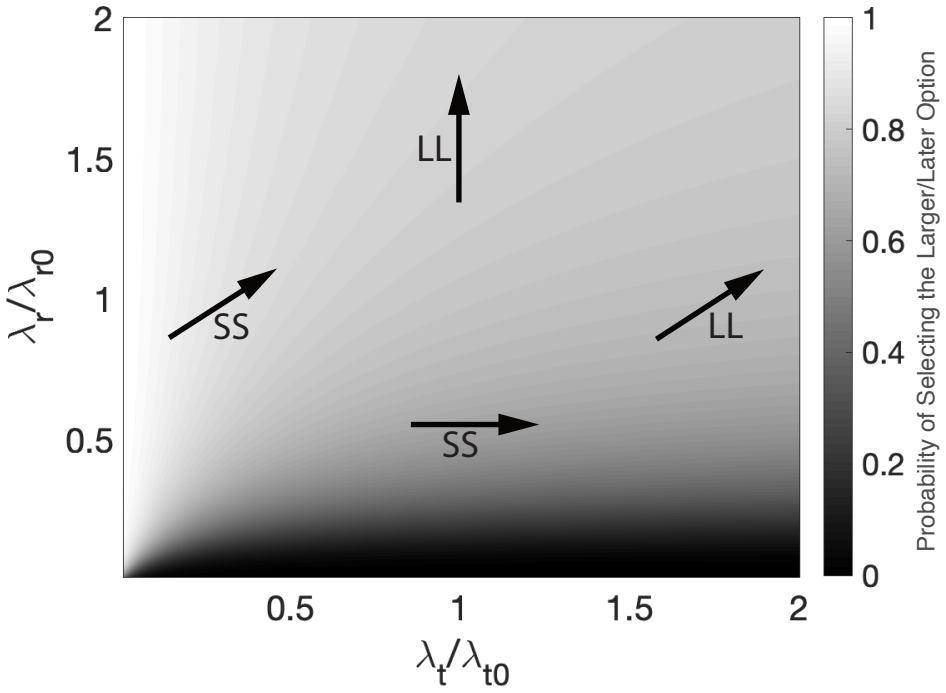

Figure 2: Behavior in ITCs depends on the relative change in reward precision compared to temporal precision. Plotted are isolines representing pairs of relative precisions that yield the same probability of selecting the larger/later option. Note that these isolines have different concavities: In the top left, the isolines are concave up (or convex), whereas in the bottom right, the isolines are concave down. Selectively increasing the reward precision promotes the larger/later option (top arrow), whereas selectively increasing the temporal precision promotes the smaller/sooner option (bottom arrow). Based on empirical findings, we assume that the temporal precision at baseline is large, compared to the baseline reward precision (each normalized by its prior precision). This means that DA's net effect is to promote the larger/later option (right arrow). If, however, the temporal precision is sufficiently reduced, DA's net effect will be to promote the smaller/sooner option (left arrow). Plotted on each axis is the ratio of encoding and prior precisions, which determines the central tendency: $w = \frac{\lambda}{\lambda+\lambda_0} = (1 + (\frac{\lambda}{\lambda_0})^{-1})^{-1}$. LL: increase in probability of selecting the larger/later option, SS: increase in probability of selecting the smaller/sooner option, λ_t: temporal encoding precision, λ_{t0}: temporal prior precision, λ_r: reward encoding precision, λ_{r0}: reward prior precision.

As discussed in the previous section, the central tendency at baseline DA levels is stronger for reward estimates than temporal estimates. It follows that the central tendency in the numerator dominates DA's influence in ITCs (Fig. 2, right arrow). Under normal conditions, then, the framework predicts that DA will promote the larger/later option, or behavior consistent with less impulsivity under higher DA levels (Fig. 3E).

This prediction matches well with empirical findings, as the majority of studies have found DA to decrease impulsivity in ITCs (15, 53, 55–60; see 22 for a recent review). For instance, Cardinal et al. (16) trained rats on an ITC involving a small reward delivered immediately and a large reward delivered after a delay that varied across blocks. After training, the authors administered DA agonists and tested the animals on the task. While the effect is smaller than in other studies (e.g., compare with Figs. 3C and 4A), the authors found that DA agonists promoted selection of the larger/later option when a visual cue was present throughout the trial (Fig. 3A).

This prediction is based on the empirically motivated result that DA's effect on the reward estimate dominates its overall effect in ITCs. However, it should be possible to elicit exactly the opposite result—an increased preference for the smaller/sooner option with DA—under conditions where the central tendency of temporal estimates dominates. For instance, timing precision has been shown to decrease when the interval salience is low (e.g., 61). Then selectively decreasing the salience during the delay period should promote the temporal central tendency and, if significant enough, overwhelm the central tendency of rewards in the numerator (Fig. 2, left arrow). Cardinal et al. (16) examined exactly this manipulation: The authors found that DA, on average, promoted selection of the larger/later option only when a salient cue was available during the delay period. If, however, a cue was present during the task *except* for the delay period, DA uncharacteristically had the opposite effect (Fig. 3B), as predicted when the temporal precision is sufficiently reduced (Fig. 3F).

It is important to note that DA manipulations can mediate preference reversals, which is captured by our model. For example, for the 20-second delay in Fig. 3A, the animal at baseline prefers the smaller/sooner option (chosen more than 50% of the time). But with high enough doses of DA agonists, it eventually comes to prefer the larger/later option (see also Figs. 3B,C,D and 4A). This empirical finding is important because it rules out hypotheses in which DA simply amplifies or mitigates existing preferences. For instance, and as mentioned above, a number of authors have proposed that DA serves a 'gain control' function on the action values during decision making (42–45). This would predict that preferences should become more extreme with higher DA levels: Preferences above the indifference (50%) line should increase, and those below the indifference line should decrease, which is inconsistent with the empirical results.

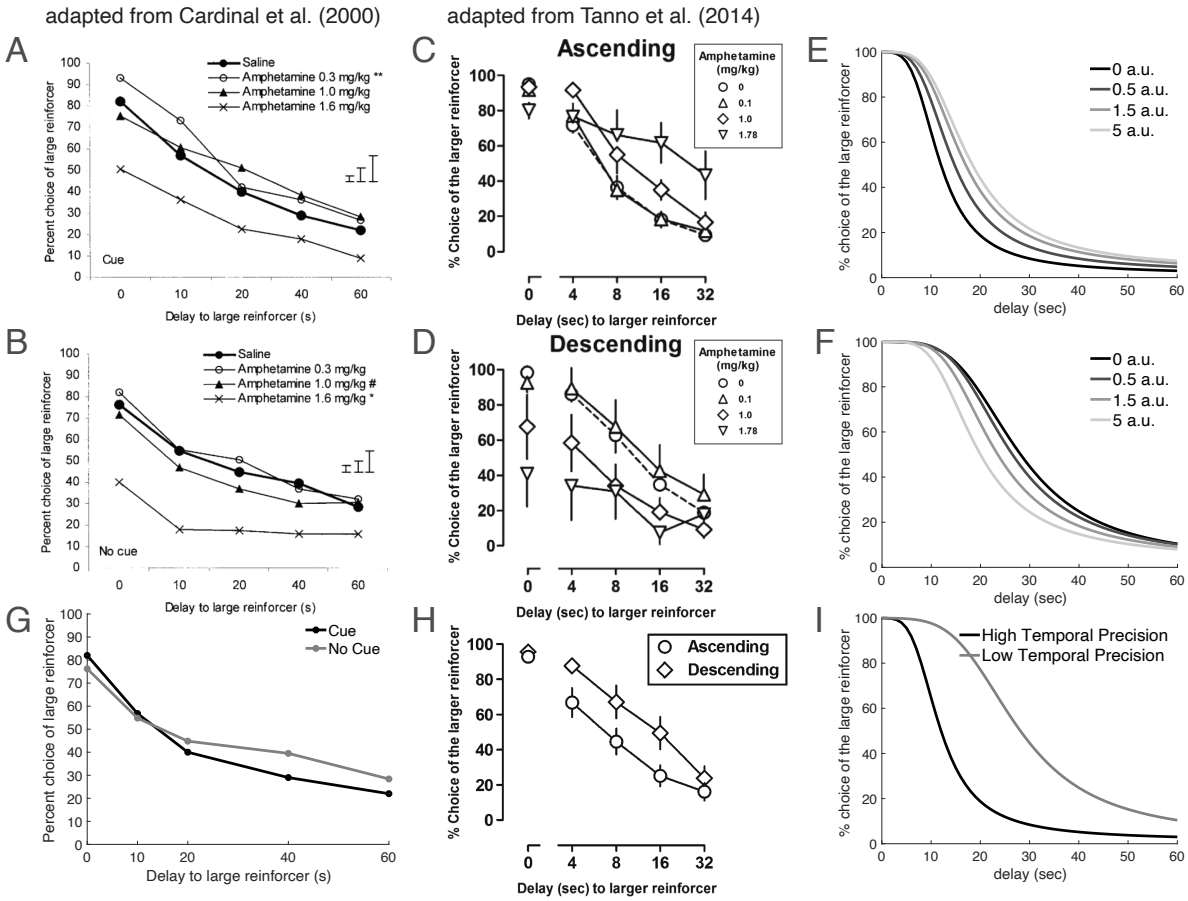

Figure 3: DA promotes selection of the larger/later option when the temporal precision is high and the smaller/sooner option when the temporal precision is low. (A) Cardinal et al. (16) trained rats on an ITC in which the animals must choose between a reward of magnitude 1 delivered immediately and a reward of magnitude 4 delivered after a delay that varied across blocks. After training, the authors administered DA agonists and examined changes in the animals' behaviors. When a cue was present during the delay period, the authors found that, with higher doses, the animals seemed less impulsive, or discounted future rewards less. This finding held for moderate changes of DA, but not the largest manipulation. (B) However, when a cue was absent during the delay period, the animals appeared more impulsive with higher doses (i.e., discounted future rewards more strongly). (C) Tanno et al. (23) administered a similar task, but varied the order in which the delays were presented. When the delays were presented in an ascending order, the rats seemed less impulsive with higher doses of DA agonists. (D) However, when the delays were presented in a descending order, the rats seemed more impulsive with higher doses. (E) Our model recapitulates these effects: Under high temporal precision, such as in the presence of a visual cue during the delay (cue condition) or as determined empirically by measuring response variability (ascending condition; $F_{1,5} = 0.11$, $p = 0.03$), DA's effect on the reward estimates will dominate in ITCs, which promotes selection of the larger/later option. (F) On the other hand, under sufficiently low temporal precision, DA's effect on the temporal estimates will dominate, which promotes selection of the smaller/sooner option. (G) At baseline, responses in the no-cue condition are biased toward the larger/later option compared to the cue condition. Note that any zero-delay difference cannot be due to a difference in the cues, since the tasks are identical in the absence of a delay. It is not clear whether these differences are statistically significant, as error bars were not provided for the saline conditions (although when the conditions were tested immediately before drug administration began, the difference was not statistically significant). Panel reproduced from the saline conditions in (A) and (B). (H) Similarly, at baseline, responses in the descending condition are biased toward the larger/later option compared to the ascending condition. (I) Our model recapitulates these effects: Selective decreases to the temporal precision promote the larger/later option. For (E, F, I), see Methods for simulation details. a.u.: arbitrary units of DA.

Though the majority of studies have found behaviors consistent with a negative correlation between impulsivity and DA, Cardinal et al. (16) found the opposite effect when the cue was selectively absent during the delay period, and we showed that the Bayesian framework captures this effect. We are aware of one other manipulation that may cause this opposite effect: In tasks where animals are trained on different delays for the larger/later option, Tanno et al. (23) have reported that DA's effect depends on the ordering of the delays. In particular, they found that DA seemed to promote choosing the larger/later option, in line with most other studies, when the delays were presented in an ascending order. However, if the delays were presented in a descending order, DA had the opposite effect (see also 62). This finding would be consistent with our framework, if the temporal precision in the ascending case were higher than that in the descending case (Fig. 3C,D). This is indeed what the authors found: When learned in an ascending order, the delay responses were less variable than when learned in a descending order. It is not clear *why* such an ordering effect exists, although one possibility is that this arises from a primacy effect in the inference about the temporal sequence, on the assumption that the initial temporal precision is higher for the short delays (e.g., 63–66).

Finally, the Bayesian framework makes a counterintuitive prediction about the relationship between baseline performance in ITCs and the effect of DA. According to our model, selectively increasing the temporal precision promotes the smaller/sooner option. However, DA's effect, when the temporal precision is already high, is to promote the *larger/later* option (compare bottom and right arrows in Fig. 2). This implies that conditions in which DA promotes the larger/later option will be conditions in which animals are, at baseline, more likely to select the smaller/sooner option. The authors of both studies above indeed observed this relationship: For both the cue and ascending conditions, animals were more likely to select the smaller/sooner option at baseline, compared to the no-cue and descending conditions, respectively (Fig. 3G,H), as predicted (Fig. 3I). Note, however, that this effect may also be due to baseline differences in the speed of the 'internal clock,' a point we turn to in the next section.

Clock speed and precision during post-reward delays

We previously mentioned the finding that temporal durations are underestimated during post-reward delays. In this section, we consider this finding more closely and examine its implications under the Bayesian framework.

There is an interesting coupling in the interval timing literature wherein DA both increases the speed of

the internal clock (67–71) and masks the central tendency effect. This relationship may be causal, as clock speed may be the mechanism through which precision is modulated (e.g., see 33). Furthermore, previous theoretical work has argued that precision will only increase when properly incentivized (72–75)—i.e., when an increase in precision improves performance. This would imply that the clock should slow down in tasks in which precision does not improve performance as well as during post-reward delays or intertrial intervals, when the animal has less control over the outcomes. These predictions have some empirical support (26, 76). Normative arguments notwithstanding, in this work it will suffice to treat the coupling between clock speed and temporal precision as an empirical phenomenon and examine its implications.

In recent primate work, Blanchard et al. (26) varied the post-reward delay in an ITC and found it to be systematically underestimated roughly by a factor of four, regardless of its total length (which varied across blocks from 0 to 10 seconds). What does a 4X reduction in clock speed imply about precision, and by extension, the central tendency? Should the presence of other post-reward delays in the same task significantly affect the animal's estimates of these delays (significant central tendency)? It is not known how exactly clock speed translates to precision, but one reasonable assumption is that the clock speed and standard deviation (inverse square root of precision) scale linearly. For instance, suppose an animal learns that it should act 8 seconds after hearing a tone, which it encodes as 8 subjective seconds (e.g., 8 ticks of the internal clock), and due to timing noise stores the interval with a granularity of 1 subjective second. This means the animal will typically respond within 7.5 and 8.5 seconds. Now imagine the internal clock were running four times slower. In that case, the animal would encode the duration as being 2 subjective seconds long (2 ticks), with a typical response occurring between 1.5 and 2.5 subjective seconds, or 6 to 10 objective (actual) seconds. Thus the standard deviation of the responses stretches by four. This in turn means that the precision will be 16 times smaller. With a large decrease in precision, our framework predicts a profound central tendency, to the point that the two posterior distributions almost overlap. Thus the animal should not discern a difference between the post-reward delays following each option. (For instance, under standard assumptions, the overlap for a 3-second and 6-second interval increases from 3% to 68%; see Methods.) On the other hand, if, as in the previous section, the salience of the post-reward delay is increased, the central tendency should become less profound, and learning the contingencies should be possible. Indeed, Blanchard et al. (26) associated each option with a different post-reward delay, and found that the animals did not learn the contingencies. Furthermore, when the authors increased the salience of the post-reward delays with a small reward at the end of each one, they found that the animals' ability to learn the option-delay contingencies was salvaged, as predicted. Note here that the posited relationship between clock speed and precision is distinct from Weber's law, which asserts that longer intervals are more noisily encoded than

shorter ones, without assuming any modifications of the clock speed (77–79, see Methods).

What does the underestimation of post-reward delays imply about behavior in ITCs? In the experiments modeled in the previous section, the authors did not impose a different post-reward delay for each option. Thus, the central tendency of the post-reward delays is not relevant, as the likelihoods, prior, and posteriors are overlapping. On the other hand, some authors have imposed different post-reward delays for each option, in order to keep the total duration for each trial constant. Thus the smaller/sooner option would have a longer post-reward delay, and the larger/later option would have a shorter post-reward delay. Notice, then, that any effect of DA on the central tendency for the post-reward delay will promote the larger/later option, which is the opposite of its effect on the pre-reward delay (what we simply referred to as the temporal duration in the previous section): With low DA, the larger/later option sees its long pre-reward delay underestimated, but its short post-reward delay overestimated, and vice versa for the smaller/sooner option. This may contribute to why DA is typically found to promote the larger/later option: Both the reward and post-reward delay have a stronger central tendency than the pre-reward delay, and for both, DA promotes the larger/later option. For instance, van Gaalen et al. (53) trained rats on an ITC in which the total duration for each trial was held constant by imposing different post-reward delays for each option. The authors found greater selection of the larger/later option with higher doses of DA agonists (Fig. 4A). Our model recapitulates this finding when it accounts for post-reward delays (Fig. 4B).

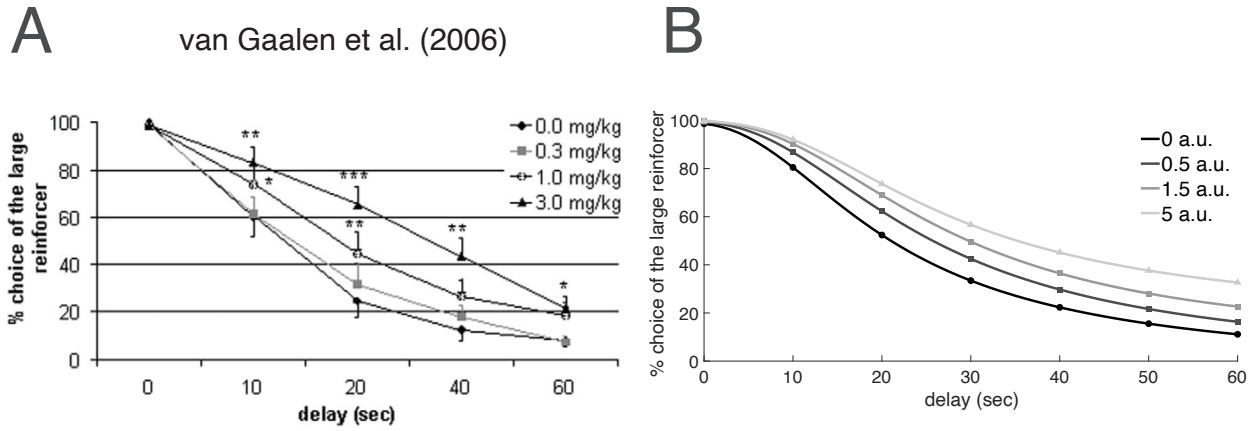

Figure 4: The central tendency of post-reward delays promotes the larger/later option. (A) van Gaalen et al. (53) trained rats on an ITC as previously described, but introduced option-specific post-reward delays to keep the total duration of each trial constant (100 seconds). By subsequently administering DA agonists, the authors found behavior consistent with less impulsivity with higher doses. Shown is the effect for the mixed DA-norepinephrine reuptake inhibitor methylphenidate. A more variable, but qualitatively similar pattern was found for the DA agonist amphetamine. (B) Our model recapitulates this effect: The interval more susceptible to the central tendency is the post-reward delay, due to its low precision, compared to the pre-reward delay. Combined with the central tendency of the reward estimates, DA's net effect is to promote the larger/later option. See Methods for simulation details. a.u.: arbitrary units of DA.

Finally, it should be noted that the coupling between clock speed and precision does not affect the DA results from the previous section. First, unlike in van Gaalen et al. (53), the post-reward delays (or intertrial intervals) in these experiments were the same for both options. Thus they are not affected by the central tendency (the likelihoods and posteriors are overlapping). Second, a faster clock amplifies temporal estimates by the gain on the clock speed. However, this amplification would only occur if the animals were trained (i.e., learned the durations) under the faster clock. Instead, the authors administered DA agonists only after the training phase.

On the other hand, the coupling does affect comparisons across the cue and no-cue conditions, and across the ascending and descending conditions. This is because, in our framework, the animal is trained on the ITCs with different temporal precisions and thus different clock speeds. Thus the trial durations during the cue and ascending conditions would be perceived to be longer than those of the no-cue and descending conditions, as they would be under the control of faster clocks. This means that, at baseline, animals in the cue and ascending conditions should favor the smaller/sooner option, as the waiting time for the larger/later option would be overestimated compared to the no-cue and descending conditions (Fig. 3I). This was indeed empirically observed, as mentioned in the previous section (Fig. 3G,H).

Correlating behavioral phenotypes

We have considered DA's effects on behaviors in interval timing and ITCs. Our final prediction, then, will be to examine how the behavioral phenomena covary with each other. Notably, we have predicted that higher DA should lead to more precise timing and lower apparent impulsivity in ITCs. Therefore, we predict that animals that are more precise timers should also appear less impulsive. Indeed, Marshall et al. (80) examined rats' impulsivity and timing abilities. To assess their impulsivity, the authors trained the rats on a standard ITC. To assess their timing precision, the authors trained the rats on a bisection task (81): Here, the rats were trained to respond with, for example, a left lever press when presented with a short (4-second) interval, and with a right lever press when presented with a long (12-second) interval. They were then tested on intermediate-duration intervals, for which they could still only respond with either a left lever press (the short-duration response) or a right lever press (the long-duration response). The stochasticity of responses was taken to reflect timing noise. The authors found that the more precise timers also tended to be less impulsive (Fig. 5A), as predicted by our framework (Fig. 5B). McClure et al. (82) also examined the correlation between timing precision and impulsivity, but using a peak-interval task, in which animals are trained to *reproduce* experienced durations, rather than a bisection task, in which animals are trained

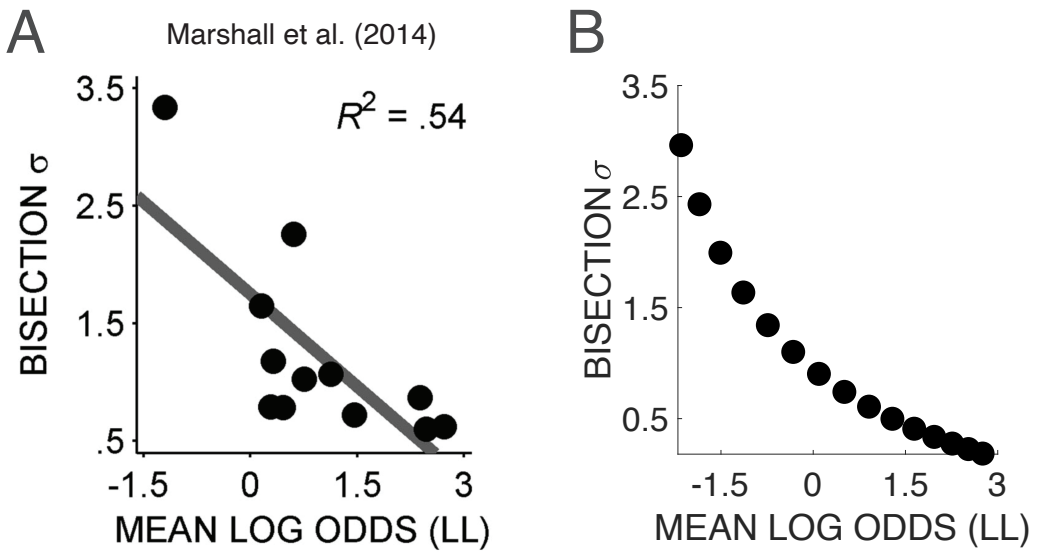

Figure 5: More precise timers are more likely to select the larger/later option. (A) Marshall et al. (80) measured rats' temporal precision and 'impulsivity' using a bisection task and ITC, respectively. The authors found that lower noise in the bisection task correlated with a tendency to select the larger/later option. LL: larger/later option, σ: parameter fit to computational model representing stochasticity of choices. (B) Our model recapitulates this effect: Animals with higher DA levels are predicted to display more precise timing and a tendency to select the larger/later option. See Methods for simulation details.

to *estimate* them, and reported similar findings.

Interestingly, Marshall et al. (80) also examined the relationship between impulsivity and reward magnitude sensitivity, which they studied using a two-armed bandit task where the larger reward was varied across blocks. The authors did not find a relationship between the two, although, as they note, this may be due to an inadequate metric for quantifying reward sensitivity (ratio of large-reward lever press rate to the sum of large- and small-reward lever press rates).

Discussion

We have shown here that DA's effects in ITCs are well-described by a Bayesian framework in which animals maximize their reward rates. Under this view, DA controls the relative influence of context in computing the reward and temporal estimates, whose ratio forms the reward rate. Most notably, the discounting-free model successfully predicts that DA should normally promote selection of the larger/later option, but should have exactly the opposite effect when the temporal precision is sufficiently low. The Bayesian view thus provides a principled framework for why DA would appear to inhibit impulsive choices in some paradigms but promote them in others.

We have followed previous theoretical and experimental work in adopting a discounting-free model of ITCs. However, our results do not necessarily rule out reward discounting more generally, nor a role for DA in this process. For instance, and as mentioned in the Introduction, humans tend to prefer smaller/sooner options even in the absence of repeated trials that make reward-rate computations meaningful. But why discount future rewards in the first place? One influential hypothesis from economics is that future rewards are discounted because of the risks involved in the delay (83). For example, a competitor may reach the reward first, or a predator may interfere in the animal's plans to collect the reward. As the delay increases, these alternative events become more likely, and the expected reward (the average over all alternatives) decreases. Another idea is that subjects respond *as if* they will have repeated opportunities to engage in the same task (84), thus mimicking the reinforcement learning problem that defines the animal variant of ITCs. More recently, Gabaix and Laibson (85) have argued that reward discounting may be due to the simulation noise involved in mentally projecting into the future: With later rewards, subjects must mentally simulate further into the future, so the simulation noise increases, and the precision decreases. Assuming a Bayesian framework with a prior centered at zero, the reward estimates will be closer to zero when rewards are more distant in the future, i.e., rewards are discounted with time (see also Gershman and Bhui (75) for an extension of this hypothesis).

Interestingly, as mentioned in the Introduction, DA seems to have the opposite effect in the human variant of the task than in the majority of animal experiments, with a promotion of the smaller/sooner option with higher DA levels. That DA may serve a qualitatively different function in the human variant is not completely unexpected, given the substantial differences in the experimental paradigms. Notably, in the human variant, (1) the subject does not actually experience the pre-reward delay, (2) there is no post-reward delay, (3) the subject does not necessarily receive an actual reward, (4) the subject may experience a single trial of this task, whereas animals are trained on many trials, and (5) the hypothetical delay is on the order of days (or months) and not seconds. Experience and repetitions may prove critical for our reinforcement learning task, and delays on the order of days engage different timing mechanisms than those on the order of seconds-to-minutes (86), which is the duration over which DA's central tendency effect has been observed. Nonetheless, the human findings may still be reconcilable with our framework under the 'repeated opportunities' hypothesis of Myerson and Green (84) mentioned above: It is possible that the temporal uncertainty surrounding durations that are not experienced, and that are on the order of days, is large and thereby dominates DA's central tendency effects. Thus DA would be predicted to promote the smaller/sooner option.

Our framework leaves open a number of theoretical and empirical questions. First, our model takes DA

to control the encoding precision, a property inherited from the Bayesian timing model of DA and further motivated by theories of DA as overcoming the cost of attention (33, 87). However, our results only require that DA control the ratio of the encoding precision to the prior precision but not necessarily the encoding precision itself. Instead, it is certainly possible that DA decreases the prior precision, as some authors have proposed (29). Interestingly, this ambiguity is not specific to theories of DA, and has been a point of debate for some Bayesian theories of autism as well (compare weak priors (88) with strong likelihoods (89)).

A second open question concerns our assumption that reward estimates are biased by a central tendency effect. Thus far, this has been inferred mainly from exploration-exploitation paradigms (see 40, for a more direct examination), but a dopaminergic modulation of reward estimates has not, to our knowledge, been observed directly. Driven by the experimental literature, we have therefore focused our simulations on manipulations of *temporal* precision. Our work then opens the door to a fruitful line of experiments with novel predictions: For instance, one can develop ITCs where the large reward is varied rather than the delay. Our framework predicts that DA will promote the larger/later option only when *reward* precision is low at baseline, and the smaller/sooner option when reward precision is high. On the other hand, selectively increasing the reward precision will always promote the larger/later option (Fig. 2). Thus, once again, by simply controlling the central tendency, DA will appear to inhibit impulsivity under some conditions, but promote it in others.

To our knowledge, this is the first framework that can accommodate the seemingly conflicting effects of DA in measures of impulsive choice, across species and experimental conditions. Nonetheless, our aim throughout this work is not to rule out a role for DA in true impulsivity, but rather to show how a single Bayesian framework can accommodate a wide range of otherwise perplexing behavioral and pharmacological phenomena.

Methods

Manipulating the precision mimics gain control of action values

Following Mikhael et al. (33), we show here that manipulating the encoding precision mimics gain control of action values.

Under standard models of action selection, the probability of selecting reward source A_i with expected reward

$\hat{\mu}_i$ follows a softmax function (90, 91):

$$p(A_i) = \frac{e^{\beta \hat{\mu}_i}}{\sum_j e^{\beta \hat{\mu}_j}}, \tag{4}$$

where β is the inverse temperature parameter, which controls choice stochasticity. A number of authors have argued that DA implements gain control on the values $\hat{\mu}_i$ in reinforcement learning tasks, possibly by controlling β (36, 44, 92). Let us then examine the case of two reward sources, A_l and A_s, yielding large and small reward, respectively. Eq. (4) can then be written as

$$p(A_l) = \frac{1}{1 + e^{-\beta(\hat{\mu}_l - \hat{\mu}_s)}}. \tag{5}$$

Notice here that the probability of exploiting the large reward depends on the difference between the reward estimates. As the quantity $\beta(\hat{\mu}_l - \hat{\mu}_s)$ increases, $p(A_l)$ increases. Hence, manipulations that increase the estimated difference will encourage exploitation, whereas manipulations that decrease it will encourage exploration. Changing the gain on the reward values (equivalent here to manipulating β) controls the influence of $(\hat{\mu}_l - \hat{\mu}_s)$ on the animal's behavior. However, this effect can also be achieved by manipulating the estimated difference $(\hat{\mu}_l - \hat{\mu}_s)$ directly. Under Bayes' rule, the encoding precision controls the resulting difference in posterior means (horizontal black segments in Fig. 1), thus mimicking gain control.

Dopamine's effect depends on its relative contribution to reward vs. temporal estimates

Our results rest on the intuition that in ITCs, DA's effect on *reward* estimation will dominate when temporal precision is high, but its effect on *temporal* estimation will dominate when temporal precision is low. We show this analytically here.

We first take the derivative of $\bar{R}$ in Eq. (3) with respect to the DA level d:

$$\frac{\partial \bar{R}}{\partial d} = \frac{\dot{w}_r(\mu_r - \mu_{r0})\hat{\mu}_t - \dot{w}_t(\mu_t - \mu_{t0})\hat{\mu}_r}{{\hat{\mu}_t}^2}, \tag{6}$$

where $\dot{w}_r = \frac{\partial w_r}{\partial d}$ and $\dot{w}_t = \frac{\partial w_t}{\partial d}$. We are interested in how DA affects $\Delta\bar{R} = \bar{R}_l - \bar{R}_s$, the difference between the larger/later reward rate estimate ($\bar{R}_l$) and the smaller/sooner reward rate estimate ($\bar{R}_s$). According to the choice rule in Eq. (5), this quantity determines the animal's behavior.

When the temporal precision is sufficiently high, $\dot{w}_t << \dot{w}_r$. Intuitively, $w_t = \frac{\lambda_t}{\lambda_t + \lambda_{t0}}$ approaches 1, so small

changes in DA do not affect it very strongly, compared to w_r. Formally, $\dot{w}_t = \frac{\dot{\lambda}_t \lambda_{t0}}{(\lambda_t+\lambda_{t0})^2}$ and $\dot{w}_r = \frac{\dot{\lambda}_r \lambda_{r0}}{(\lambda_r+\lambda_{r0})^2}$, where $\dot{\lambda}_t = \frac{\partial \lambda_t}{\partial d}$ and $\dot{\lambda}_r = \frac{\partial \lambda_r}{\partial d}$. Because the prior precisions are finite, we require that $\frac{\dot{\lambda}_t}{\dot{\lambda}_r} << \frac{\lambda_t^2}{\lambda_r^2}$, so that $\dot{w}_t << \dot{w}_r$.

It follows that, in Eq. (6), the first term in the numerator dominates:

$$\frac{\partial \bar{R}}{\partial d} \simeq \frac{\dot{w}_r(\mu_r - \mu_{r0})}{\hat{\mu_t}}. \tag{7}$$

The term in the parentheses is positive for the larger/later option and negative for the smaller/sooner option. Then, $\frac{\partial \bar{R}_l}{\partial d} > 0$ and $\frac{\partial \bar{R}_s}{\partial d} < 0$. It follows that $\frac{\partial \Delta \bar{R}}{\partial d} > 0$, so DA promotes the larger/later option.

Similarly, when the temporal precision is sufficiently low, $\dot{w}_t >> \dot{w}_r$. Formally, we require that $\frac{\dot{\lambda}_t}{\dot{\lambda}_r} >> \frac{\lambda_{r0}\lambda_{t0}}{(\lambda_r+\lambda_{r0})^2}$, so that small changes in DA strongly affect λ_t and, by extension, w_t. In this case, the second term in the numerator of Eq. (6) dominates:

$$\frac{\partial \bar{R}}{\partial d} \simeq -\frac{\dot{w}_t(\mu_t - \mu_{t0})\hat{\mu}_r}{\hat{\mu_t}^2}. \tag{8}$$

The term in the parentheses is positive for the larger/later option and negative for the smaller/sooner option. Then, $\frac{\partial \bar{R}_l}{\partial d} < 0$ and $\frac{\partial \bar{R}_s}{\partial d} > 0$. It follows that $\frac{\partial \Delta \bar{R}}{\partial d} < 0$, so DA promotes the smaller/sooner option.

Note here that the approximation in Eq. (8) depends on the durations being different for each option. Otherwise, $(\mu_t - \mu_{t0}) = 0$, and the first term in the numerator in Eq. (6) will always dominate, regardless of how low the temporal precision is. In this case, DA will always promote the larger option. Said differently, if the delays are equal, the task reduces to a simple two-armed bandit task (the options are equivalent except for a difference in reward magnitudes), and our framework predicts that DA will always promote the larger option.

The Bayesian framework preserves the hyperbolic relationship with delays

A well-replicated result is that animals behave as though discounting future rewards hyperbolically or near-hyperbolically (21, 54). The hypothesis that animals seek to maximize their reward rates in ITCs preserves this empirical phenomenon. Here, the animal's choice is determined by:

$$\bar{R} = \frac{r}{PRE + POST}, \tag{9}$$

where $\bar{R}$ is the reward rate, r is the reward, PRE is the pre-reward delay, and $POST$ is the post-reward delay (including the intertrial interval). The reward rate has a hyperbolic relationship with the pre-reward delay. An important concern, then, is whether the Bayesian framework preserves this relationship. We show here that it does.

First, parametrizing over the delays while holding the reward magnitudes constant, we can rewrite Eq. (3) as

$$\bar{R} = \frac{A}{w_t \mu_t + (1 - w_t)\mu_{t0} + POST}, \tag{10}$$

where A is a constant. Because the prior is determined by the distribution of stimuli in the context, its mean μ_{t0} and standard deviation s_{t0} scale approximately linearly with μ_t (roughly, μ_{t0} is the average of μ_{ts}, which is small and fixed, and μ_{tl}, which we parametrize over). So we can further write:

$$\bar{R} \simeq \frac{A}{w_t \mu_t + (1 - w_t)(B\mu_t) + POST} \tag{11}$$

$$= \frac{A}{(w_t(1 - B) + B)\mu_t + POST}, \tag{12}$$

where B is a constant. Notice that the discount factor (term in outer parentheses in second line) increases with the temporal precision. This is exactly our result from Fig. 3I.

The hyperbolic relationship will only hold if w_t is also a constant (thus making the discount factor constant). We can write this as:

$$w_t = \frac{\lambda_t}{\lambda_t + \lambda_{t0}} \tag{13}$$

$$= \frac{s_{t0}^2}{s_t^2 + s_{t0}^2}, \tag{14}$$

where $s_t = \frac{1}{\sqrt{\lambda_t}}$ and $s_{t0} = \frac{1}{\sqrt{\lambda_{t0}}}$ are the standard deviations of the likelihood and prior, respectively.

We assume that the standard deviation s_t changes in accordance with Weber's law: $s_t = \alpha\mu_t$, where α is known as the Weber fraction (77–79), and, as stated above, s_{t0} scales approximately linearly with the longer duration: $s_{t0} \simeq \beta\mu_t$, where both α and β are constants. Then,

$$w_t \simeq \frac{\beta^2}{\alpha^2 + \beta^2}, \tag{15}$$

which is a constant, as required. Thus, as a result of Weber's law, Eq. (10) describes a hyperbolic relationship between the delay and the reward rate.

It is interesting that the hyperbolic relationship requires a linearity between the likelihood standard deviation and the delay, which is exactly the empirically observed Weber's law. Because of Weber's law, the Bayesian framework is characterized by scale invariance: Amplifying the temporal intervals in a task does not affect the extent of each interval's central tendency. It is worth speculating whether this serves an evolutionary purpose.

Simulation details

For our results to hold, we require that the mapping between DA and encoding precision be monotonic over the relevant domain. Weber's law applies in the domains of reward magnitudes (93–95) and interval timing (77–79); therefore, it will be convenient to refer to the encoding standard deviation $s = \frac{1}{\sqrt{\lambda}}$. We arbitrarily set $s(d) = \frac{\alpha\mu+\epsilon}{d}$, where α is the Weber fraction, ϵ represents signal-independent noise, and d is the DA level. When the DA level is fixed to 1, this relation reduces to the generalized Weber's law (e.g., 96). We treat a DA level of 1 as the baseline (e.g., saline) condition. Increasing the DA level d decreases the standard deviation s and thus increases the precision λ, as required. We set α and ϵ to 0.4 and 0.5, respectively for rewards, and to 0.15 and 1, respectively for timing.

We have assumed in the Results that the central tendency is more profound in the domain of rewards than in the domain of timing under normal conditions (high temporal precision). This is achieved by setting the Weber fraction to be higher for rewards. For conditions in which the temporal precision is selectively and sufficiently reduced, we increase the encoding standard deviation by a factor of 8. Note that our choice of $\alpha = 0.15$ is a typical Weber fraction for rodents in interval timing (97).

The prior mean and standard deviation were set to the mean and standard deviation of the distribution of stimuli. The prior precision λ_0 is an inverse function of the prior variance σ_0^2, $\lambda_0 = \frac{1}{1+\sigma_0^2}$, where the added '1' in the denominator is so the precision does not go to infinity when the two stimuli are equal (a form of Laplace smoothing).

For Figs. 3 and 4, we set the reward magnitudes for the smaller/sooner and larger/later options to 1 and 4, respectively, in accordance with the number of pellets used as reinforcement in both Cardinal et al. (16) and van Gaalen et al. (53). The post-reward delay was arbitrarily fixed to 50 in Fig. 3, whereas in Fig. 4, the post-reward delay was 100 minus the pre-reward delay (total trial length was fixed to 100 in van Gaalen et al. (53)). We assumed the post-reward delays were underestimated by a factor of 8. As per the Results, we assumed that the encoding standard deviation increases by the same factor (here, 8), compared to the

encoding standard deviation for pre-reward delays. Finally, the agent makes decisions based on the softmax choice rule in Eq. (4) with $\beta = 25$.

For Fig. 5, the DA level was varied between 0.4 and 1.6 to mimic natural differences across animals, while being centered at 1. DA levels were sampled logarithmically between these two extremes. For the ITC, and in accordance with the experimental setup of Marshall et al. (80), the reward magnitudes were 1 and 2, and the post-reward delay was 120. The short duration was 2.5, 5, 10, or 30, and the long duration was always 30. The mean log odds were computed by averaging over the log odds for each temporal pair. We assumed the post-reward delay was underestimated by a factor of 4. An inverse temperature parameter of $\beta = 500$ was required to match the data well, although the mismatch between this β value and that of the previous experiments may in part be due to the arbitrarily defined effect of DA on the encoding precision (in Eq. (5), β is multiplied by the posterior mean difference, whose relationship with DA is monotonic, but arbitrarily set). All other parameters are identical to those used in the experiments above. Finally, for the bisection task, the probability of selecting the 'long' response was the probability of the long duration for each time point, divided by the sum of probabilities of the short and long durations. The stochasticity parameter σ was fit to the softmax function in Eq. (4), where $\sigma = \beta^{-1}$.

Effect of encoding precision on learning post-reward delays

Under standard conditions, a 4X increase in the encoding standard deviation can result in a profound increase in the overlap between the posterior distributions for a short interval μ_s and a long interval μ_l. It will be convenient to consider the posterior standard deviations $\hat{s} = \frac{1}{\sqrt{\lambda+\lambda_0}}$. The overlap can be computed by first identifying the time $t_c \in [\mu_s, \mu_l]$ at which the posterior probabilities are equal (intersection of the two distributions):

$$t_c = \frac{\hat{\mu}_s \hat{s}_l^2 - \hat{s}_s \left(\hat{\mu}_l \hat{s}_s + \hat{s}_l \sqrt{(\hat{\mu}_l - \hat{\mu}_s)^2 + 2(\hat{s}_l^2 - \hat{s}_s^2) \ln(\frac{\hat{s}_l}{\hat{s}_s})} \right)}{\hat{s}_l^2 - \hat{s}_s^2}. \tag{16}$$

Then the overlap is

$$p(\mu_s, \mu_l) = \int_{-\infty}^{t_c} \mathcal{N}(t; \hat{\mu}_l, \hat{s}_l^2) dt + \int_{t_c}^{+\infty} \mathcal{N}(t; \hat{\mu}_s, \hat{s}_s^2) dt. \tag{17}$$

Intuitively, this represents the sum of the area under both curves to the left and to the right of t_c, respectively.

As above, we take the Weber fraction to be 0.15. Plugging in, it follows that the overlap between the posterior distributions for a 3-second interval and a 6-second interval is 0.03 (area under the curves; maximum is 1) but increases to 0.68 when the Weber fraction is 0.15 x 4 = 0.6.

Acknowledgements

The authors are grateful to Rahul Bhui for comments on an earlier draft of the paper.

Funding

The project described was supported by National Institutes of Health grants T32GM007753 (JGM), T32MH020017 (JGM), and U19 NS113201-01 (SJG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

J.G.M. and S.J.G. developed the model and contributed to the writing of the paper. J.G.M. analyzed and simulated the model, made the figures, and wrote the first draft.

Competing interests

The authors declare no competing interests.

Data and code availability

Source code for all simulations can be found at www.github.com/jgmikhael/impulsivity.

References

[1] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[2] Rosemary Tannock, Russell J Schachar, Robert P Carr, Diane Chajczyk, and Gordon D Logan. Effects of methylphenidate on inhibitory control in hyperactive children. *Journal of abnormal child psychology*, 17(5):473–491, 1989.

[3] Christopher Gillberg, Hans Melander, Anne-Liis von Knorring, Lars-Olof Janols, Gunilla Thernlund, Bruno Hägglöf, Lena Eidevall-Wallin, Peik Gustafsson, and Svenny Kopp. Long-term stimulant treatment of children with attention-deficit hyperactivity disorder symptoms: a randomized, double-blind, placebo-controlled trial. *Archives of general psychiatry*, 54(9):857–864, 1997.

[4] Robert L Findling and Judith W Dogin. Psychopharmacology of ADHD: children and adolescents. *The Journal of clinical psychiatry*, 59:42–49, 1998.

[5] Mary V Solanto. Neuropsychopharmacological mechanisms of stimulant drug action in attention-deficit hyperactivity disorder: a review and integration. *Behavioural brain research*, 94(1):127–152, 1998.

[6] Keri Shiels, Larry W Hawk Jr, Brady Reynolds, Rebecca J Mazzullo, Jessica D Rhodes, William E Pelham Jr, James G Waxmonsky, and Brian P Gangloff. Effects of methylphenidate on discounting of delayed rewards in attention deficit/hyperactivity disorder. *Experimental and clinical psychopharmacology*, 17(5):291, 2009.

[7] Erin A Heerey, Benjamin M Robinson, Robert P McMahon, and James M Gold. Delay discounting in schizophrenia. *Cognitive neuropsychiatry*, 12(3):213–221, 2007.

[8] James M Gold, James A Waltz, Kristen J Prentice, Sarah E Morris, and Erin A Heerey. Reward processing in schizophrenia: a deficit in the representation of value. *Schizophrenia bulletin*, 34(5): 835–847, 2008.

[9] Nora D Volkow, Joanna S Fowler, Gene-Jack Wang, James M Swanson, and Frank Telang. Dopamine in drug abuse and addiction: results of imaging studies and treatment implications. *Archives of neurology*, 64(11):1575–1579, 2007.

[10] Warren K Bickel, David P Jarmolowicz, E Terry Mueller, Mikhail N Koffarnus, and Kirstin M Gatchalian. Excessive discounting of delayed reinforcers as a trans-disease process contributing to

addiction and other disease-related vulnerabilities: emerging evidence. *Pharmacology & therapeutics*, 134(3):287–297, 2012.

[11] Alain Dagher and Trevor W Robbins. Personality, addiction, dopamine: insights from Parkinson's disease. *Neuron*, 61(4):502–510, 2009.

[12] Sean S O'Sullivan, Andrew H Evans, and Andrew J Lees. Dopamine dysregulation syndrome. *CNS drugs*, 23(2):157–170, 2009.

[13] Harriet de Wit, Justin L Enggasser, and Jerry B Richards. Acute administration of d-amphetamine decreases impulsivity in healthy volunteers. *Neuropsychopharmacology*, 27(5):813–825, 2002.

[14] Alex Pine, Tamara Shiner, Ben Seymour, and Raymond J Dolan. Dopamine, time, and impulsivity in humans. *Journal of Neuroscience*, 30(26):8888–8896, 2010.

[15] Tammy R Wade, Harriet de Wit, and Jerry B Richards. Effects of dopaminergic drugs on delayed reward as a measure of impulsive behavior in rats. *Psychopharmacology*, 150(1):90–101, 2000.

[16] Rudolf N Cardinal, Trevor W Robbins, and Barry J Everitt. The effects of d-amphetamine, chlordiazepoxide, α-flupenthixol and behavioural manipulations on choice of signalled and unsignalled delayed reinforcement in rats. *Psychopharmacology*, 152(4):362–375, 2000.

[17] Jeffrey R Stevens and David W Stephens. The adaptive nature of impulsivity. 2010.

[18] Howard Rachlin and Leonard Green. Commitment, choice and self-control 1. *Journal of the experimental analysis of behavior*, 17(1):15–22, 1972.

[19] George Ainslie. Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological bulletin*, 82(4):463, 1975.

[20] Henry Tobin and Alexandra W Logue. Self-control across species (Columba livia, Homo sapiens, and Rattus norvegicus). *Journal of Comparative Psychology*, 108(2):126, 1994.

[21] Howard Rachlin. *The science of self-control*. Harvard University Press, 2000.

[22] Valérie D'Amour-Horvat and Marco Leyton. Impulsive actions and choices in laboratory animals and humans: effects of high vs. low dopamine states produced by systemic treatments given to neurologically intact subjects. *Frontiers in behavioral neuroscience*, 8:432, 2014.

[23] Takayuki Tanno, David R Maguire, Cedric Henson, and Charles P France. Effects of amphetamine and methylphenidate on delay discounting in rats: interactions with order of delay presentation. *Psychopharmacology*, 231(1):85–95, 2014.

[24] Alex Kacelnik. Normative and descriptive models of decision making: time discounting and risk sensitivity. In *CIBA foundation symposium*, pages 51–70. Wiley Online Library, 1997.

[25] Nathaniel D Daw and David S Touretzky. Behavioral considerations suggest an average reward TD model of the dopamine system. *Neurocomputing*, 32:679–684, 2000.

[26] Tommy C Blanchard, John M Pearson, and Benjamin Y Hayden. Postreward delays and systematic biases in measures of animal temporal discounting. *Proceedings of the National Academy of Sciences*, 110(38):15491–15496, 2013.

[27] Vijay Mohan K Namboodiri, Stefan Mihalas, Tanya Marton, and Marshall Gilmer Hussain Shuler. A general theory of intertemporal decision-making and the perception of time. *Frontiers in Behavioral Neuroscience*, 8:61, 2014.

[28] Karl J Friston, Tamara Shiner, Thomas FitzGerald, Joseph M Galea, Rick Adams, Harriet Brown, Raymond J Dolan, Rosalyn Moran, Klaas Enno Stephan, and Sven Bestmann. Dopamine, affordance and active inference. *PLoS Computational Biology*, 8(1):e1002327, 2012.

[29] Vincent D Costa, Valery L Tran, Janita Turchi, and Bruno B Averbeck. Reversal learning and dopamine: a Bayesian perspective. *Journal of Neuroscience*, 35(6):2407–2416, 2015.

[30] Chara Malapani, Brian Rakitin, R Levy, Warren H Meck, Bernard Deweer, Bruno Dubois, and John Gibbon. Coupled temporal memories in Parkinson's disease: a dopamine-related dysfunction. *Journal of Cognitive Neuroscience*, 10(3):316–331, 1998.

[31] Chara Malapani, Bernard Deweer, and John Gibbon. Separating storage from retrieval dysfunction of temporal memory in Parkinson's disease. *Journal of Cognitive Neuroscience*, 14(2):311–322, 2002.

[32] Zhuanghua Shi, Russell M Church, and Warren H Meck. Bayesian optimization of time perception. *Trends in Cognitive Sciences*, 17(11):556–564, 2013.

[33] John G Mikhael, Lucy Lai, and Samuel J Gershman. Rational inattention and tonic dopamine. *bioRxiv*, 2020. doi: 10.1101/2020.10.04.325175.

[34] Christoph Eisenegger, Michael Naef, Anke Linssen, Luke Clark, Praveen K Gandamaneni, Ulrich Müller, and Trevor W Robbins. Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology*, 39(10):2366, 2014.

[35] Eunjeong Lee, Moonsang Seo, Olga Dal Monte, and Bruno B Averbeck. 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*, 35(16):6298–6306, 2015.

[36] François Cinotti, Virginie Fresno, Nassim Aklil, Etienne Coutureau, Benoît Girard, Alain R Marchand, and Mehdi Khamassi. Dopamine blockade impairs the exploration-exploitation trade-off in rats. *Scientific reports*, 9(1):6770, 2019.

[37] Jeff A Beeler, Nathaniel D Daw, Cristianne RM Frazier, and Xiaoxi Zhuang. Tonic dopamine modulates exploitation of reward learning. *Frontiers in behavioral neuroscience*, 4:170, 2010.

[38] Arif A Hamid, Jeffrey R Pettibone, Omar S Mabrouk, Vaughn L Hetrick, Robert Schmidt, Caitlin M Vander Weele, Robert T Kennedy, Brandon J Aragona, and Joshua D Berke. Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, 19:117–126, 2016.

[39] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4):292–303, 2008.

[40] Samuel J Gershman and Yael Niv. Novelty and inductive generalization in human reinforcement learning. *Topics in cognitive science*, 7(3):391–415, 2015.

[41] Hrvoje Stojić, Eric Schulz, Pantelis P Analytis, and Maarten Speekenbrink. It's new, but is it good? How generalization and uncertainty guide the exploration of novel options. *Journal of Experimental Psychology: General*, 2020.

[42] Jonathan D Cohen and David Servan-Schreiber. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychological review*, 99(1):45, 1992.

[43] Todd S Braver, Jonathan D Cohen, and David Servan-Schreiber. A computational model of prefrontal cortex function. In *Advances in neural information processing systems*, pages 141–148, 1995.

[44] Anne GE Collins and Michael J Frank. Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological review*, 121(3):337, 2014.

[45] Bruno B Averbeck and Vincent D Costa. Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience*, 20(4):505, 2017.

[46] Mehrdad Jazayeri and Michael N Shadlen. Temporal context calibrates interval timing. *Nature neuroscience*, 13(8):1020, 2010.

[47] Luigi Acerbi, Daniel M Wolpert, and Sethu Vijayakumar. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS computational biology*, 8(11):e1002771, 2012.

[48] Karin M Bausenhart, Oliver Dyjas, and Rolf Ulrich. Temporal reproductions are influenced by an internal reference: Explaining the Vierordt effect. *Acta Psychologica*, 147:60–67, 2014.

[49] Katja M Mayer, Massimiliano Di Luca, and Marc O Ernst. Duration perception in crossmodally-defined intervals. *Acta psychologica*, 147:2–9, 2014.

[50] Neil W Roach, Paul V McGraw, David J Whitaker, and James Heron. Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences*, 114(2): 412–417, 2017.

[51] Benjamin J De Corte and Matthew S Matell. Temporal averaging across multiple response options: insight into the mechanisms underlying integration. *Animal cognition*, 19(2):329–342, 2016.

[52] Mathias Pessiglione, Ben Seymour, Guillaume Flandin, Raymond J Dolan, and Chris D Frith. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442 (7106):1042, 2006.

[53] Marcel M van Gaalen, Reinout van Koten, Anton NM Schoffelmeer, and Louk JMJ Vanderschuren. Critical involvement of dopaminergic neurotransmission in impulsive decision making. *Biological psychiatry*, 60(1):66–73, 2006.

[54] James E Mazur. An adjusting procedure for studying delayed reinforcement. *Commons, ML.; Mazur, JE.; Nevin, JA*, pages 55–73, 1987.

[55] Catharine A Winstanley, Jeffrey W Dalley, David EH Theobald, and Trevor W Robbins. Global 5-HT depletion attenuates the ability of amphetamine to decrease impulsive choice on a delay-discounting task in rats. *Psychopharmacology*, 170(3):320–331, 2003.

[56] F Denk, ME Walton, KA Jennings, T Sharp, MFS Rushworth, and DM Bannerman. Differential

involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology*, 179(3):587–596, 2005.

[57] Catharine A Winstanley, Quincey LaPlant, David EH Theobald, Thomas A Green, Ryan K Bachtell, Linda I Perrotti, Ralph J DiLeone, Scott J Russo, William J Garth, David W Self, et al. ΔFosB induction in orbitofrontal cortex mediates tolerance to cocaine-induced cognitive dysfunction. *Journal of Neuroscience*, 27(39):10497–10507, 2007.

[58] Stan B Floresco, TL Maric, and Sarvin Ghods-Sharifi. Dopaminergic and glutamatergic regulation of effort-and delay-based decision making. *Neuropsychopharmacology*, 33(8):1966–1979, 2008.

[59] Mikhail N Koffarnus, Amy H Newman, Peter Grundt, Kenner C Rice, and James H Woods. Effects of selective dopaminergic compounds on a delay discounting task. *Behavioural pharmacology*, 22(4):300, 2011.

[60] Neil E Paterson, Caitlin Wetzler, Adrian Hackett, and Taleen Hanania. Impulsive action and impulsive choice are mediated by distinct neuropharmacological substrates in rat. *International Journal of Neuropsychopharmacology*, 15(10):1473–1487, 2012.

[61] Qingqing Li, Peiduo Liu, Shunhang Huang, and Xiting Huang. The effect of phasic alertness on temporal precision. *Attention, Perception, & Psychophysics*, 80(1):262–274, 2018.

[62] David R Maguire, Cedric Henson, and Charles P France. Effects of amphetamine on delay discounting in rats depend upon the manner in which delay is varied. *Neuropharmacology*, 87:173–179, 2014.

[63] Adam N Sanborn, Thomas L Griffiths, and Daniel J Navarro. Rational approximations to rational models: alternative algorithms for category learning. *Psychological review*, 117(4):1144, 2010.

[64] Roger P Levy, Florencia Reali, and Thomas L Griffiths. Modeling the effects of memory on human online sentence processing with particle filters. In *Advances in neural information processing systems*, pages 937–944, 2009.

[65] Joshua T Abbott and Thomas L Griffiths. Exploring the influence of particle filter parameters on order effects in causal learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 33, 2011.

[66] Pratiksha Thaker, Joshua B Tenenbaum, and Samuel J Gershman. Online learning of symbolic concepts. *Journal of Mathematical Psychology*, 77:10–20, 2017.

[67] Andres V Maricq, Seth Roberts, and Russell M Church. Methamphetamine and time estimation. *Journal of Experimental Psychology: Animal Behavior Processes*, 7(1):18, 1981.

[68] Andres V Maricq and Russell M Church. The differential effects of haloperidol and methamphetamine on time estimation in the rat. *Psychopharmacology*, 79(1):10–15, 1983.

[69] Warren H Meck. Affinity for the dopamine D2 receptor predicts neuroleptic potency in decreasing the speed of an internal clock. *Pharmacology Biochemistry and Behavior*, 25(6):1185–1189, 1986.

[70] Ruey-Kuang Cheng, Yusuf M Ali, and Warren H Meck. 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*, 88(2):149–159, 2007.

[71] Jessica I Lake and Warren H Meck. Differential effects of amphetamine and haloperidol on temporal reproduction: dopaminergic regulation of attention and clock speed. *Neuropsychologia*, 51(2):284–292, 2013.

[72] Christopher A Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3): 665–690, 2003.

[73] Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203, 2015.

[74] Filip Matějka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98, 2015.

[75] Samuel J Gershman and Rahul Bhui. Rationally inattentive intertemporal choice. *Nature communications*, 11(1):1–8, 2020.

[76] Lewis A Bizo and K Geoffrey White. Reinforcement context and pacemaker rate in the behavioral theory of timing. *Learning & behavior*, 23(4):376–382, 1995.

[77] John Gibbon. Scalar expectancy theory and Weber's law in animal timing. *Psychological review*, 84(3): 279, 1977.

[78] Russell M Church and W Meck. A concise introduction to scalar timing theory. *Functional and neural mechanisms of interval timing*, pages 3–22, 2003.

[79] JER Staddon. Some properties of spaced responding in pigeons. *Journal of the Experimental Analysis of Behavior*, 8(1):19–28, 1965.

[80] Andrew T Marshall, Aaron P Smith, and Kimberly Kirkpatrick. Mechanisms of impulsive choice: I. individual differences in interval timing and reward processing. *Journal of the Experimental Analysis of Behavior*, 102(1):86–101, 2014.

[81] Russell M Church and Marvin Z Deluty. Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavior Processes*, 3(3):216, 1977.

[82] Jesse McClure, Jeffrey Podos, and Heather N Richardson. Isolating the delay component of impulsive choice in adolescent rats. *Frontiers in integrative neuroscience*, 8:3, 2014.

[83] Paul A Samuelson. A note on measurement of utility. *The review of economic studies*, 4(2):155–161, 1937.

[84] Joel Myerson and Leonard Green. Discounting of delayed rewards: Models of individual choice. *Journal of the experimental analysis of behavior*, 64(3):263–276, 1995.

[85] Xavier Gabaix and David Laibson. Myopia and discounting. Technical report, National bureau of economic research, 2017.

[86] Catalin V Buhusi and Warren H Meck. What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, 6(10):755–765, 2005.

[87] Sanjay G Manohar, Trevor T-J Chong, Matthew AJ Apps, Amit Batla, Maria Stamelou, Paul R Jarman, Kailash P Bhatia, and Masud Husain. Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology*, 25(13):1707–1716, 2015.

[88] Elizabeth Pellicano and David Burr. When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in cognitive sciences*, 16(10):504–510, 2012.

[89] Rebecca P Lawson, Geraint Rees, and Karl J Friston. An aberrant precision account of autism. *Frontiers in human neuroscience*, 8:302, 2014.

[90] Roger N Shepard. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, 55(6):509, 1958.

[91] R. Duncan Luce. *Individual Choice Behavior: a Theoretical Analysis*. John Wiley and sons, 1959.

762 [92] Mark D Humphries, Mehdi Khamassi, and Kevin Gurney. Dopaminergic control of the exploration-
763 exploitation trade-off via the basal ganglia. *Frontiers in neuroscience*, 6:9, 2012.

764 [93] Peter R Killeen, Heather Cate, and Trung Tran. Scaling pigeons'choice of feeds: Bigger is better.
765 *Journal of the Experimental Analysis of Behavior*, 60(1):203–217, 1993.

766 [94] Melissa Bateson and Alex Kacelnik. Accuracy of memory for amount in the foraging starling, sturnus
767 vulgaris. *Animal Behaviour*, 50(2):431–443, 1995.

768 [95] Alex Kacelnik and Melissa Bateson. Risky theories—the effects of variance on foraging decisions. *Amer-*
769 *ican Zoologist*, 36(4):402–434, 1996.

770 [96] David J Getty. Discrimination of short temporal intervals: A comparison of two models. *Perception &*
771 *Psychophysics*, 18(1):1–8, 1975.

772 [97] CR Gallistel, Adam King, and Robert McDonald. Sources of variability and systematic error in mouse
773 timing behavior. *Journal of Experimental Psychology: Animal Behavior Processes*, 30(1):3, 2004.