When to choose: Information seeking in the speed-accuracy tradeoff

Javier Masís*
Princeton Neuroscience Institute
Princeton University
Princeton, NJ, USA
jmasis@princeton.edu

David E. Melnikoff*
Department of Psychology
Northeastern University
Boston, MA, USA
davidemelnikoff@gmail.com

Lisa Feldman Barrett
Department of Psychology
Northeastern University
Boston, MA, USA
l.barrett@northeastern.edu

Jonathan D. Cohen
Princeton Neuroscience Institute
Princeton University
Princeton, NJ, USA
jdc@princeton.edu

Abstract

Normative accounts of decision-making predict that people attempt to balance the immediate rewards associated with correct responses against the costs of deliberation. However, humans frequently deliberate longer than normative models say they should. We propose that people try to optimize not only their rate of material rewards, but also their rate of information gain. A computational model that implements this idea successfully mimics human decision makers, reproducing key patterns of behavior not predicted by alternative models. Moreover, simulations reveal a normative basis for our model: An agent that exchanges even a small amount of immediate reward for information will improve its decision-making ability through learning, allowing it to earn more reward in the long run than an agent disinterested in information. Maximizing a combination of reward and information rate is a simple yet effective strategy for solving the speed-accuracy tradeoff that may resolve lingering mysteries about human decision-making.

Keywords: learning; decision-making; drift-diffusion model; optimality; reward maximization; information theory

Introduction

Normative models of decision-making predict that humans try to maximize the rate at which they are rewarded by optimizing how long they spend deliberating: long enough to make informed decisions, but not so long as to waste precious time (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012; Gold & Shadlen, 2002). Mysteriously, this prediction is often violated. Humans systematically fail to maximize their reward rate by spending too much time deliberating (Bogacz, Hu, Holmes, & Cohen, 2010; Balci et al., 2011).

Overly long deliberation has been attributed to two main sources (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Zacksenhouse, Bogacz, & Holmes, 2010): a preference for accuracy (operationalized as an *error penalty*), and imperfect estimates of temporal delays between trials (operationalized as *temporal uncertainty*). However, these factors fail to fully explain the phenomenon of overly long deliberation. They erroneously predict that the tendency to spend too long deliberating vanishes when the decision at hand is particularly difficult. The intuition behind this result is that decision-makers—even those who care about accuracy and represent temporal uncertainty—will spend minimal time deliberating if the correct response seems impossible to divine. As sensible as it would seem, this behavior is not observed empirically. On the contrary, it is precisely when choices are very difficult that overly long deliberation is most pronounced (Balci et al., 2011; Holmes & Cohen, 2014; Masís, Chapman, Rhee, Cox, & Saxe, 2023).

To solve this puzzle, a new model is needed that hews closer to the empirical data. We propose such a model by drawing on evidence that humans and non-human animals are driven to accumulate not just rewards, but also *information* (Bromberg-Margin & Hikosaka, 2009; Litovsky, Loewenstein, Horn, & Olivola, 2022; Bennett, Bode, Brydevall, Warren, & Murawski, 2022; Cogliati Dezza, Schulz, & Wu, 2022; Melnikoff, Carlson, & Stillman, 2022). We suggest that when determining how long to deliberate, decision-makers have a default tendency to optimize not their rate of reward, but their rate of information gain. A model that implements this idea mimics the human tendency to deliberate too long regardless of the difficulty of the choice at hand. Thus, the basic drive to consume information may play a crucial role in the speed-accuracy trade-off.

Beyond better capturing people's behavior, why might decision-makers prioritize information over reward in the first place? According to recent theorizing, overly slow responding may promote learning, allowing decision-makers to improve their performance and, ultimately, accrue more reward in the long run (Masís, Musslick, & Cohen, 2021). So, by encouraging overly slow responding, information seeking may yield more *cumulative* reward through the improvement of decision-making performance. In line with this idea, we find that when our model is endowed with the ability to learn, its cumulative reward increases the more it prioritizes information. This result supports our model from a normative standpoint as well as an empirical one. Specifically, it matches behavior recently observed in rodents (Masís et al., 2023): Much like our model, rats unlock higher future rewards by exchanging immediate rewards for the learning benefits of overly slow responding. Our model shows how this seemingly sophisticated, forward-looking process can be achieved through the simple, myopic policy of maximizing one's current rate of information gain.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

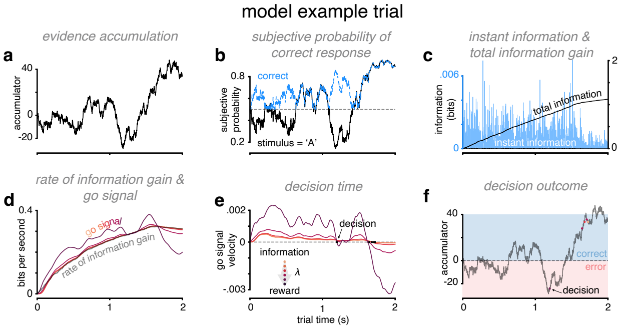

Figure 1: **Model description.** **(a)** Accumulator state. Positive values correspond to evidence in favor of stimulus 'A' (the true stimulus) and negative values correspond to evidence in favor of stimulus 'B'. **(b)** $p(\text{stimulus} = \text{'A'})_k$ (black line) and subjective probability of correct response $\pi(\text{correct})_k$ (blue line). **(c)** Instantaneous information i_k (blue lines) via eq. 6, and total information $\sum_{j=0}^{k} i_j$ (black line). **(d)** Rate of information gain y_k (black line) via eq. 7 and go signal (warm tone lines, with reward priority λ increasing as colors darken) via eq. 8. **(e)** A decision is made when the go signal x_k reaches a local maximum, i.e. its velocity is $v_k \leq 0$ (black dots denote this moment for various values of λ). **(f)** Accumulator state at the decision time determines the trial outcome.

Model

Background

In a standard drift-diffusion model (DDM), decision-makers accumulate noisy evidence over time (Ratcliff & McKoon, 2008). Given a set of prior beliefs about the evidence-generating process, decision-makers can use the accumulated evidence to compute, at each moment in time, the posterior probability that a given option is correct (Calder-Travis, Bogacz, & Yeung, 2020; Drugowitsch et al., 2012; Moran, 2015). A decision is made once the accumulated evidence reaches a predetermined bound (Ratcliff & McKoon, 2008), which denotes how confident a decision-maker needs to be about which option is correct before making a decision. This bound can be constant or change over time (Drugowitsch et al., 2012; Moran, 2015), and its location will impact the agent's reward rate. There exists an optimal bound for every signal-to-noise ratio (SNR) and inter-trial interval that can be parametrized as a relation between decision time (DT) and error rate (ER) called the optimal performance curve (OPC) (Bogacz et al., 2006). Overly slow responding is defined as any response that lies above the OPC.

Model Overview

We propose that decision-makers have a default tendency to maximize their rate of information gain, but can strategically increase their rate of reward. In our model, this behavior is implemented through the computation of a "go signal" at each time step k of deliberation. A decision is made when the go signal reaches a local maximum. The trajectory of the go signal depends on two things. First, it depends on the decision-maker's rate of information gain, which changes throughout deliberation. Second, it depends on whether, and to what extent, the decision-maker prioritizes reward rate. If the decision-maker assigns zero priority to reward rate, the go signal and the information rate will reach local maximums at approximately the same time, resulting in a decision that promotes information seeking. However, to the extent that the decision-maker does prioritize its reward rate, the go signal will be biased to plateau at a closer to reward-rate-optimal time step.

We present our model in three parts. First, we describe the basic process through which the go signal evolves over time. Second, we describe how the rate of information gain is computed and used to guide the trajectory of the go signal. Finally, we describe how the trajectory of the go signal can be biased in order to increase the decision-maker's rate of reward.

Model Description

The Go Signal. Δt is a small time interval between each step of the decision-making process. Over the course of this time interval, the go signal changes at a constant velocity. Accordingly, the projected value of the go signal at time step k is given by

$$\hat{x}_k = x_{k-1} + \Delta t v_{k-1} \tag{1}$$

where $\hat{x}_k$ is the projection, x_{k-1} is the go signal at the previous time step, and v_{k-1} is the velocity of the go signal at the previous time step. The projected velocity $\hat{v}_k$ is simply

$$\hat{v}_k = v_{k-1} \tag{2}$$

The go signal evolves in a closed loop manner. At each time step, its projected value $\hat{x}_k$ is compared to a target value y_k, and the residual (i.e., the difference between the projection and the target value) r_k is used to adjust the go signal and its velocity in such a way as to minimize the residual at the next time step. r_k is computed according to

$$r_k = y_k - \hat{x}_k \tag{3}$$

and is used to compute the new go signal x_k and its velocity v_k as follows:

$$x_k = \hat{x}_k + \alpha r_k \tag{4}$$

$$v_k = \hat{v}_k + \frac{\beta}{\Delta t} r_k \tag{5}$$

where $0 < \alpha < 1$ and $0 < \beta \leq 2$ control the amount of adjustment applied to the projections. Equations (1)–(5) are equivalent to an α–β filter and, for certain values of α and β, a steady-state Kalman filter. They ensure that the trajectory of the go signal hews closely to that of the target value by reducing the magnitude of the residual at each time step. In addition, when $0 < \beta < 1$, noise in the go signal is suppressed, making its trajectory smoother than that of the target value. The suppression of noise is critical, since the decision-maker's policy is to make a decision when the go signal reaches a local maximum; random noise may produce "false maximums," causing the decision-maker to choose well before a "true maximum" occurs in the underlying signal. In all simulations, we set the initial state values $x_{k=0} = v_{k=0} = 0$.

Rate of Information Gain. The target value toward which the go signal is adjusted is the decision-maker's rate of information gain, computed as follows.

Faced with a binary forced-choice, a decision-maker accumulates evidence according to a standard DDM (Ratcliff & McKoon, 2008) (Fig. 1a). Assuming a Bayesian observer, the accumulator state (i.e., the total evidence accumulated) is used to compute, at every time step, the decision-maker's "confidence," denoted as c_k. Confidence is the log odds of the decision-maker choosing correctly at time k assuming a "greedy" response policy of always choosing the most probable option (Calder-Travis et al., 2020; Drugowitsch et al., 2012; Moran, 2015) .

c_k defines the probability distribution π_k, over decision outcomes (correct vs. incorrect) (Fig. 1b), which is used to compute information gain at each time step:

$$i_k = D_{\mathrm{KL}}[\pi_k || \pi_{k-1}] \tag{6}$$

D_{KL} is Kullback-Leibler divergence, which, in this case, quantifies the relative entropy from π_{k-1} to π_k. Therefore, i_k quantifies the amount of information that the decision-maker's kth accumulator state provides about its probability of responding correctly (Fig. 1c, blue). In other words, i_k tells the decision-maker the marginal information benefit from having accumulated evidence for one more time step.[1]

At each time step, the agent uses $i_{0:k}$ to track its rate of information gain y_k—that is, the total information gained (Fig. 1c, black) divided by time spent deliberating k, non-decision time t_0, and the mean response-to-stimulus interval (RSI). We express this value in units of bits-per-second (bps) (Fig. 1d, grey):

$$y_k = \frac{\sum_{j=0}^{k} i_j}{k + t_0 + \mathrm{RSI}} \tag{7}$$

y_k is used to compute x_k and v_k according to (4) and (5), respectively. With a small β, this results in a smooth go signal that tracks the trajectory of the decision-maker's rate of information gain. Accordingly, a decision-maker that makes its choice when the go signal reaches a local maximum will come close to maximizing its information rate. The reason the decision-maker optimizes its go signal, as opposed to its information rate directly, is because the decision-maker may wish to optimize a combination of information rate and reward rate. The decision-maker can accomplish this by optimizing the go signal, which is sensitive to both rates, as we will now show.

Rate of Reward. Our model allows decision-makers to bias the go signal by increasing or decreasing its velocity. In other words, the decision-maker can accelerate or decelerate the go signal. This entails the following modifications to (1) and (2). The projected go signal at time step k becomes

$$\hat{x}_k = x_{k-1} + \Delta t v_{k-1} + \tfrac{1}{2}\Delta t^2 a_{k-1} \tag{8}$$

and its projected velocity becomes

$$\hat{v}_k = v_{k-1} + \Delta t a_{k-1} \tag{9}$$

where a_{k-1} is the acceleration applied by the decision-maker at time step $k-1$.

The acceleration is a function of two quantities. First, it is a function of the degree to which the decision-maker prioritizes reward $\lambda \in \mathbb{R}_{\geq 0}$. Second, it is a function of the difference between the decision-maker's current level of confidence c_k and its optimal confidence threshold c_{op}. The optimal confidence threshold is the value of c_k at which the decision-maker must make its choice in order to maximize its reward rate. c_{op} can be computed numerically for any given SNR and inter-trial interval (Bogacz et al., 2006). Acceleration at time step k is computed according to:

$$a_k = \lambda(c_{\mathrm{op}} - c_k) \tag{10}$$

According to (10), when $c_{\mathrm{op}} > c_k$, amplifying reward priority λ will accelerate the go state. This, in turn, will make the go state less likely to reach a local maximum, preventing the decision-maker from making its choice prior to the time step at which it would maximize its reward rate. Conversely, when $c_{\mathrm{op}} < c_k$, amplifying reward priority will decelerate the go signal, which will make it more likely to reach a local maximum, thereby preventing the decision-maker from deliberating longer than it should to maximize its rate of reward. In the limit $\lambda \to \infty$ the decision-maker will optimize only its reward rate. When $\lambda = 0$, the decision-maker will optimize only its rate of information gain. For all other values of λ the decision-maker will optimize some combination of information rate and reward rate, with greater emphasis on reward rate as λ increases.

Summary. The go signal evolves in a closed loop fashion. Its value and velocity are adjusted at each time step to reduce the difference between its trajectory and the trajectory of the decision-maker's information rate. In addition, the decision-maker can prioritize reward rate by accelerating or decelerating the go signal, which will encourage it to plateau closer to the reward-rate-maximizing time step. In our model simulations, a decision is made when the velocity of the go signal becomes non-positive. The response is considered correct if, at that time step, the accumulator state favors the correct answer. Otherwise, the response is considered incorrect. We set $\alpha = .0248$ and $\beta = .0003$. These settings apply sufficient noise suppression to produce smooth go signal trajectories, and render our model equivalent to a steady-state Kalman filter with process noise $\sigma_p^2 = .0001$ and observation

[1] D_{KL} is defined so long as the probability of a particular decision outcome never reaches 1 or 0—a condition that is always satisfied due to the noise in the accumulation process.

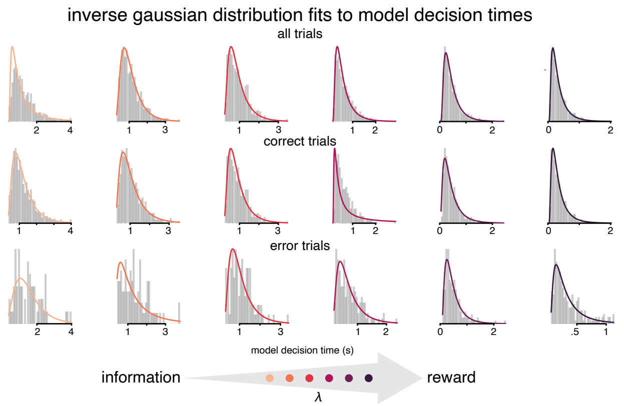

Figure 2: **Information seeking produces realistic response times.** Inverse Gaussian fits to model decision times. Warm-tone colors denote information to reward priority (increasing λ parameter). Top, middle, and bottom row show all, correct, and error trials respectively

noise $\sigma_o^2 = 1000$. The reward priority λ was always held constant throughout the decision-making process. For our simulations, we compare the performance of our model to that of a standard DDM agent set to optimize its instantaneous rate of reward (iRR)—that is, an iRR-optimal policy—, as well as agents with an error penalty and with temporal uncertainty.

Results

Information Seeking Produces Realistic Response Times

Any plausible model of decision-making must produce realistic distributions of response times. To confirm that our model clears this initial hurdle, we used it to simulate 1000 response times across a range of λ values, and used these data to fit Inverse Gaussian (also known as Wald) distributions, which descriptively capture the response time distributions of human and non-human decision-makers (Luce et al., 1986; Matzke & Wagenmakers, 2009). We consistently obtained reasonable fits, confirming that our model produces empirically valid distributions of response times (Fig. 2).

Information Seeking Produces Overly Slow Responding Across all Levels of Difficulty

There are two main alternative models of overly slow responding (Bogacz et al., 2006; Zacksenhouse et al., 2010; Holmes & Cohen, 2014). The *error penalty* model introduces a negative utility q for errors, leading subjects to display accuracy-bias-like behavior:

$$\mathrm{iRR}_{EP} = \frac{1 - ER - qER}{DT + t_0 + \mathrm{RSI}} \tag{11}$$

The *temporal uncertainty* model assumes subjects have a noisy estimate of the intertrial intervals, with some proportionality constant a capturing the presumed level of uncertainty:

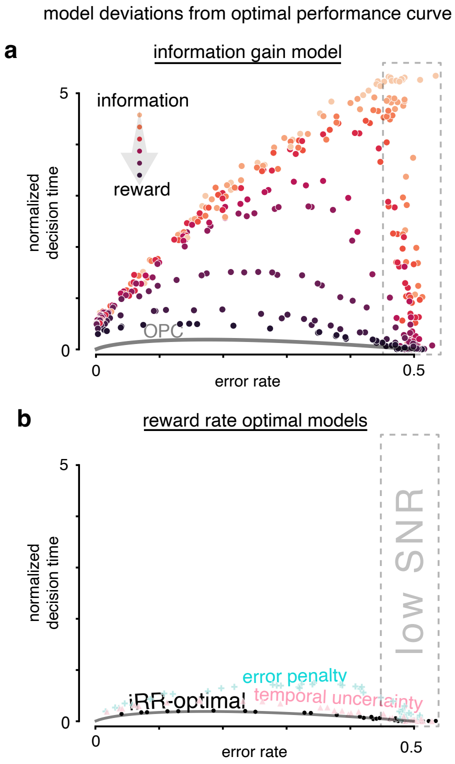

Figure 3: **Model performance in speed-accuracy space. (a)** information gain model ER & DT as functions of SNR (dots), and reward priority λ (warm tones). OPC in grey. Overly slow responding (i.e., distance above OPC), is amplified with greater priority on information rate. At low SNR (dashed box), model produces large range of response times. **(b)** iRR-optimal (black), error penalty (light blue), temporal uncertainty (pink) optimal models. At low SNR, optimal models converge to OPC.

$$\mathrm{iRR}_{TU} = \frac{1 - ER}{DT + (1+a)(t_0 + \mathrm{RSI})} \tag{12}$$

Thresholds can then be optimized according to these reward rate equations, resulting in alternative optimal performance curves (Zacksenhouse et al., 2010).

Our model mimics the behavior of non-human animals and people (Masís et al., 2023; Balci et al., 2011; Bogacz et al., 2010; Zacksenhouse et al., 2010) by spending longer deliberating that it should to maximize its reward rate (Fig. 3a). This was the case across all parameter settings, but especially for small values of λ (when information rate is prioritized most). Critically, unlike the alternative models (error penalty, and temporal uncertainty, Fig. 3b), our model responds too slowly even when the decision at hand is so difficult that accuracy is at chance. These results support our hypothesis that a decision policy that maximizes a combination of reward rate and information rate reproduces key patterns of human decision-making that existing models on their own cannot capture.

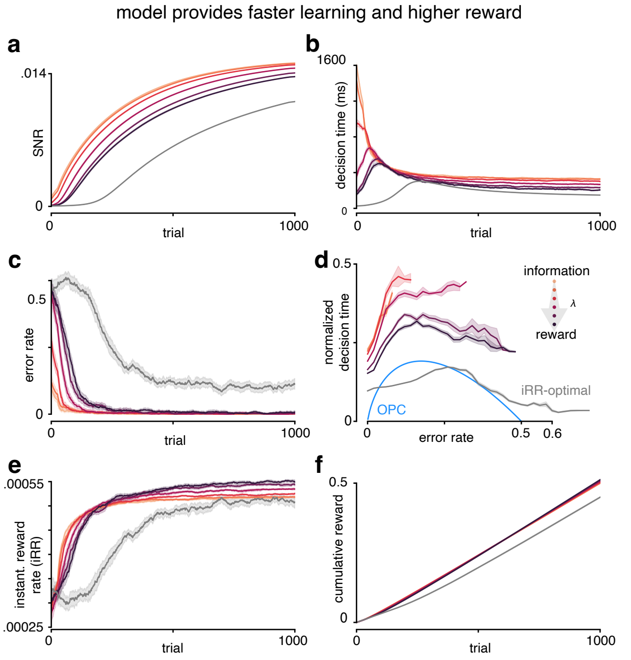

Figure 4: **Information seeking yields faster learning and higher reward. (a,b,c,e,f)** SNR, ER, DT, iRR and cumulative reward over trials, separated by reward priority λ (warm tones) and iRR-optimal policy (grey). **(d)** Trajectory in speed-accuracy space during learning.

Information Seeking Yields Faster Learning and Higher Reward

To investigate the potential benefits of a decision policy that prioritizes information, we endow our model with the ability to improve its decision-making performance through learning. Recent modeling work shows that longer deliberation times lead to faster learning (Masís et al., 2023). Thus, to capture learning, we assume that SNR improves as a function of deliberation time; the longer an agent deliberates, the more they are able to learn (i.e., reduce SNR) and, in turn, improve their future performance. Under this assumption, our model learns faster than an iRR-optimal agent endowed with the same learning abilities (Fig. 4a). In fact, even a small regard for information (large λ) yields considerable learning benefits over pure reward rate maximization (Fig. 4a,b; dark purple vs. grey).

This stark difference arises because, as noted above, our model spends more time deliberating, with the difference in deliberation being more pronounced as a decreasing function of λ—that is, as information is prioritized over reward (Fig. 4b). Notably, this difference in deliberation is highest at the start of learning when it is (fortuitously) most beneficial for improvement, and lowest as the agent's skill level reaches mastery (ER $\approx$ 0).

One would expect an information seeking policy to result in reward rate opportunity costs, and indeed this is the case. Our model initially accrues less reward than an iRR-optimal agent (Fig. 4e). However, this difference is surprisingly small when noting the considerable differences in decision time behavior (Fig. 4b). Despite this initial reward opportunity cost, across a broad swath of reward priority values λ, our model earns more cumulative reward over learning than an iRR-optimal agent (Fig. 4f). We note that at high skill levels (resulting in low ERs and low DTs), the differences in behavior between information seeking and reward-maximizing agents shrink. This observation is important because it demonstrates that the reward opportunity costs for information seeking experienced at low skill levels are compensated for with the learning benefit that comes precisely with information seeking. Thus, although it would be preferable to maximize reward rate in the narrow case of a completely non-stationary environment, in the more realistic and much broader set of cases where learning is possible because there is periodic stationarity, information seeking becomes preferable.

These findings suggest that our model has a normative basis. In addition, they align with the vast literature on automaticity and skill learning (Newell & Rosenbloom, 1981), and recent empirical results: rats, like our model, have been found to outperform an iRR-optimal policy in terms of cumulative reward by making overly slow responses on early decision-making trials, and speeding up as their learning plateaus (Masís et al., 2023).

Discussion

We have proposed a novel account of a pervasive and perplexing finding: human and non-human decision-makers systematically deliberate longer than they should to optimize their rate of reward. On our account, this (apparent) suboptimality arises because decision-makers wish to maximize not only their reward rate, but also their information rate. A model embodying this hypothesis confirms that information seeking can produce overly slow responding. Critically, our model produces overly slow responding at all levels of difficulty—even at difficulty levels where decision-makers respond with chance accuracy—which is a commonly observed phenomenon that alternative models miss.

Beyond empirical support, our model enjoys normative support: When learning is taken into account, it improves its performance faster and, ultimately, accrues more rewards. Maximizing a mix of reward and information rates is a relatively simple and myopic strategy that solves the difficult intertemporal choice problem of how to weigh present rewards against information that can help with future rewards. Previous work has proposed solving this problem via cognitive control, assigning more cognitive control to prioritize information over reward when the expectation to learn is high (Masís et al., 2021). This model, built upon the Expected Value of Control (EVC) theory (Shenhav, Botvinick, & Cohen, 2013), formulates the previously described intertemporal choice problem across trials, and as such requires a prediction of future discounted reward. The model presented herein formulates the problem within a trial, resulting in desirable behavior across trials, without the need for a prediction of future states. In the absence of metacognitive control, our model thus provides a candidate heuristic, or default algorithm, that solves this problem naturally and myopically. Modulating the reward priority λ (through metacognitive control) could serve as a proxy for expected learning prospects, and as such could be annealed over time with experience.

Our model raises the question of whether the speed-accuracy trade-off can be reduced to a trade-off between reward seeking and information seeking. Indeed, according to our model, the only mechanism through which a decision-maker can modulate its rate of responding is by changing its reward priority λ—that is, the degree to which it prioritizes the maximization of information rate relative to reward rate. This is implausible, however, because it implies that decisions-makers are incapable of responding faster than reward-rate optimal. Our model's decision-making speed is maximized in the limit $\lambda \to \infty$, at which point it becomes a pure optimizer of reward rate. It seems unlikely that human decision-makers are incapable of faster-than-optimal responding. This suggests that decision-makers attempt to optimize not just reward rate and information rate, but other variables as well—some of which favor particularly fast responding. Identifying these additional variables will be an important direction for future work.

Acknowledgments

J.M. is supported by the Presidential Postdoctoral Research Fellowship at Princeton University, and by the National Institutes of Health institutional training grant T32MH065214. D.E.M. is supported by the National Institute of Mental Health of the National Institutes of Health under award number F32MH124430.

References

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: reward rate ultimately beats accuracy. *Attention, Perception, & Psychophysics*, *73*(2), 640–657.

Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2022). Intrinsic valuation of information in decision making under uncertainty. *PLOS Computational Biology*, *12*(7), e1005020.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700.

Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed–accuracy trade-off that maximizes reward rate? *The Quarterly Journal of Experimental Psychology*, *63*(5), 863–891.

Bromberg-Margin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, *63*(1), 119–126.

Calder-Travis, J., Bogacz, R., & Yeung, N. (2020). Bayesian confidence for drift diffusion observers in dynamic stimuli tasks. *BioRxiv*.

Cogliati Dezza, I., Schulz, E., & Wu, C. M. (2022). *The drive for knowledge: The science of human information-seeking*. Cambridge: Cambridge University Press.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, *32*(11), 3612–3628.

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*(2), 299–308.

Holmes, P., & Cohen, J. D. (2014). Optimality and some of its discontents: Successes and shortcomings of existing models for binary decisions. *Topics in cognitive science*, *6*(2), 258–278.

Litovsky, Y., Loewenstein, G., Horn, S., & Olivola, C. Y. (2022). Loss aversion, the endowment effect, and gain-loss framing shape preferences for noninstrumental information. *Proceedings of the National Academy of Sciences*, *119*(34), e2202700119.

Luce, R. D., et al. (1986). *Response times: Their role in inferring elementary mental organization* (No. 8). Oxford University Press on Demand.

Masís, J., Chapman, T., Rhee, J. Y., Cox, D. D., & Saxe, A. M. (2023). Strategically managing learning during perceptual decision making. *Elife*, *12*, e64978.

Masís, J. A., Musslick, S., & Cohen, J. (2021). The value of learning and cognitive control allocation. In *Proceedings of the annual meeting of the cognitive science society.*

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychonomic bulletin & review*, *16*(5), 798–817.

Melnikoff, D. E., Carlson, R. W., & Stillman, P. E. (2022). A computational theory of the subjective experince of flow. *Nature Communications*, *3*(1), 1–13.

Moran, R. (2015). Optimal decision making in heterogeneous and biased environments. *Psychonomic bulletin & review*, *22*(1), 38–53.

Newell, A., & Rosenbloom, P. S. (1981). Cognitive skills and their acquisition. In J. R. Anderson (Ed.), (chap. Mechanisms of Skill Acquisition and the Law of Practice). Hillsdale, NJ: Erlbaum.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217-240.

Zacksenhouse, M., Bogacz, R., & Holmes, P. (2010). Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology*, *54*(2), 230–246.