

# Confluence of timing and reward biases in perceptual decision-making dynamics

Maxwell Shinn<sup>1,3</sup>, Daniel Ehrlich<sup>1,3</sup>, Daeyeol Lee<sup>2†</sup>, John D. Murray<sup>1,3†</sup>, Hyojung Seo<sup>1,3†</sup>

**\*For correspondence:**

[maxwell.shinn@yale.edu](mailto:maxwell.shinn@yale.edu) (MS);  
[john.murray@yale.edu](mailto:john.murray@yale.edu) (JDM);  
[hyojung.seo@yale.edu](mailto:hyojung.seo@yale.edu) (HS)

<sup>†</sup>These authors contributed equally to this work

<sup>1</sup>Department of Psychiatry, New Haven, CT, USA; <sup>2</sup>Department of Neuroscience, Yale University, New Haven, CT, USA; <sup>3</sup>Interdepartmental Neuroscience Program, Yale University, New Haven, CT, USA

**Present address:** <sup>†</sup>Johns Hopkins University, Baltimore, MD, USA

## Abstract

Although the decisions of our daily lives often occur in the context of temporal and reward structures, the impact of such regularities on decision-making strategy is poorly understood. Here, to explore how temporal and reward context modulate strategy, we trained rhesus monkeys to perform a novel perceptual decision-making task with asymmetric rewards and time-varying evidence reliability. To model the choice and response time patterns, we developed a computational framework for fitting generalized drift-diffusion models (GDDMs) which flexibly accommodates diverse evidence accumulation strategies. We found that a dynamic urgency signal and leaky integration, in combination with two independent forms of reward biases, best capture behavior. We also tested how temporal structure influences urgency by systematically manipulating the temporal structure of sensory evidence, and found that the time course of urgency was affected by temporal context. Overall, our approach identified key components of cognitive mechanisms for incorporating temporal and reward structure into decisions.

## Introduction

In an uncertain and dynamic environment, humans and other animals detect temporal regularities in the environment and use them to inform decision-making (*Woodrow, 1914; Rosenbaum and Collyer, 1998; Los, 2010; Behrens et al., 2007; Farashahi et al., 2017*). Even in relatively simple perceptual tasks, the timing and accuracy of decisions are sensitive to the temporal statistics of stimuli (*Grosjean et al., 2001*). Cognitive strategies in perceptual tasks are often reflected in the response times (RTs), which therefore can be effectively used to test mechanistic models of decision-making processes (*Luce, 1986*).

One paradigm for studying perceptual decision-making computations and their neural correlates is to present dynamic sensory evidence over time (*Newsome et al., 1989; Roitman and Shadlen, 2002*). Evidence accumulation has been proposed as a leading strategy for decision-making under this paradigm, which can be formalized using the drift-diffusion model (DDM) (*Ratcliff, 1978; Ratcliff et al., 2016*). The DDM has been employed to capture choice and RT behavior in a range of decision-making tasks (*Palmer et al., 2005; Krajbich and Rangel, 2011; Ding and Gold, 2012; Bogacz et al., 2006; Resulaj et al., 2009; Ratcliff et al., 2003; Mormann et al., 2010*). In the DDM, a dynamic decision variable integrates evidence over time, and a decision is reached when this variable crosses a bound. An open question is how evidence accumulation is shaped by tem-

poral uncertainty or expectation of the stimulus. For instance, if the signal-to-noise ratio changes across the time of stimulus presentation, can the integration process dynamically gate this input signal?

Such gating is important for integrating relevant information while ignoring irrelevant information. Uncertainty in the temporal onset of information creates a need to modulate the integration process over time. For example, beginning integration before relevant information is available causes noise to be unnecessarily integrated, thereby reducing decision accuracy (Laming, 1979; Grosjean et al., 2001). By contrast, beginning integration after the onset of the stimulus discards potentially beneficial information (Devine et al., 2019). It is unknown how these timing factors should be implemented within a DDM framework. Time-varying bounds (Murphy et al., 2016; Hanks et al., 2014; Ditterich, 2006b; Drugowitsch et al., 2012) and gain modulation (Ditterich, 2006b; Cisek et al., 2009; Thura et al., 2012) have improved predictive ability for RT patterns, but these mechanisms have not been utilized to explain uncertainty in stimulus onset time. Furthermore, the utility of DDMs incorporating these and other complex extensions is limited by the computational challenges of simulating and fitting such models to empirical choice and RT patterns.

In addition to sensory evidence and temporal expectation, decisions should incorporate differences in value among the options. Reward bias has long been an active topic of research (Feng et al., 2009; Rorie et al., 2010; Laming, 1968; Ratcliff, 1985; Mulder et al., 2012), with more recent attention to its connections to timing (Drugowitsch et al., 2012; Voskuilen et al., 2016; Hanks et al., 2011; Lauwereyns et al., 2002; Nagano-Saito et al., 2012; Gao et al., 2011). While these asymmetries have previously been incorporated into the DDM framework (Ratcliff and McKoon, 2008; Ratcliff, 1985; Edwards, 1965; Laming, 1968; Mulder et al., 2012), the strategies through which sensory, value, and timing information are combined to form a decision are still poorly understood.

To investigate these issues, we developed a novel behavioral paradigm for perceptual decision-making in which the onset of the evidence is temporally uncertain. Therefore, the temporal structure of evidence could be strategically exploited during decision-making. By simultaneously manipulating reward differences between options, we also examined whether and how time-varying evidence interacts with temporal strategies to produce reward bias. We trained monkeys to perform this task, and found that the animals exploited the temporal structure of the evidence and adjusted the timing of their decisions in such a way to reduce uncertainty about the evidence onset.

To quantitatively model the decision-making behavior, we developed a framework for computationally efficient simulation and fitting of generalized DDMs (GDDMs) to choice and RT behavior. GDDMs encompass both the DDM and also non-integrative models as special cases (Cisek et al., 2009; Thura et al., 2012). We found that a GDDM with timing and reward processes can quantitatively capture the animal's behavior as determined by its RT distribution. Temporal modulation of the integration process can be implemented through an urgency signal which varies across time. Such an urgency signal can flexibly adjust the animal's behavior as task timing demands change. In addition, capturing reward-related behavioral effects required two mechanisms, one dependent on the integration process and the other independent of integration. Overall, our findings suggest that evidence accumulation in perceptual decision-making can be flexibly modulated by temporal and reward expectation, and that generalized DDMs can mathematically capture these phenomena.

## Results

### Behavioral task

Two rhesus monkeys (*Macaca mulatta*) were trained to perform a two-alternative forced choice, color matching task (Figure 1). In each trial, a central square patch was presented consisting of a 20x20 grid of green and blue pixels that rearranged randomly at 20 Hz. Stimulus presentation was divided into two consecutive periods containing an uninformative "presample" and informative "sample". The animal indicated its choice by shifting its gaze to one of two flanking choice

90 targets, one green and one blue. The trial was rewarded via juice delivery if the selected target  
 91 color corresponded to the majority color of pixels in the sample. Task difficulty was manipulated  
 92 by parametrically varying the proportion of pixels of each color in the sample, which we refer to  
 93 as color coherence. Animals were allowed to direct their gaze to a choice target any time after  
 94 the onset of the sample. In addition, reward cues were displayed surrounding the saccade targets  
 95 which indicated whether a large or small reward would be delivered for a correct response to the  
 96 corresponding target. These reward cues were randomly assigned to a target on each trial. Zero  
 97 coherence corresponds to a 50/50 mixture of blue and green pixels, and coherence of  $\pm 1$  corre-  
 98 sponds to a solid color. By convention, we define coherence relative to large- and small-reward  
 99 targets, where a coherence of +1 or -1 corresponds to all pixels being the color of the large-reward  
 100 or small-reward target, respectively.

101 Temporal uncertainty in the sensory evidence was introduced by varying the duration of the  
 102 uninformative presample. No explicit cue was presented to indicate this transition from presample  
 103 to sample. During the presample period, the color coherence was zero, and at the transition to  
 104 the sample, the coherence switched to the particular value chosen for that trial. The presample  
 105 duration was selected randomly from three possible time intervals—0, 0.4 or 0.8 s—with equal  
 106 probability. A premature choice was punished by a 2 s timeout. In a separate experiment described  
 107 below, we manipulated the set of presample durations (*Figure 6*).

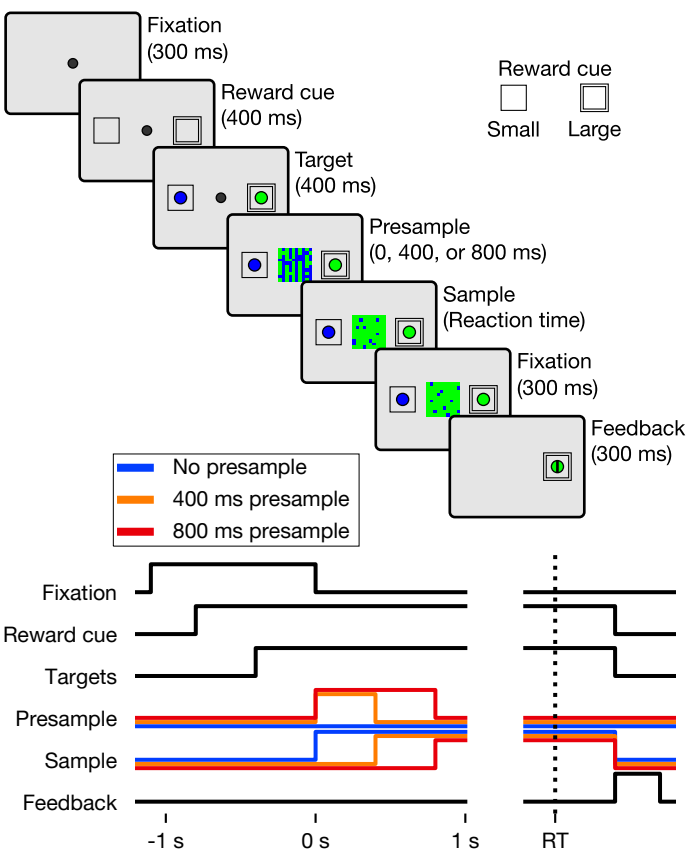
# 108 Task parameters influence behavior

109 We predicted that all three primary task variables—coherence, presample duration, and reward  
 110 location—would lead to systematic changes in the animal's behavior as measured by the RT dis-  
 111 tributions. In this study, RT was measured from the onset of the presample, since the transition  
 112 from presample to sample was not explicitly cued. To test the impact of these three task variables  
 113 on the animal's performance, we used multiple-regression models to predict RT and accuracy in  
 114 individual trials using these three task variables (see Methods).

115 We found that all three task variables had a significant influence on RT ( $p < .05$  for both ani-  
 116 mals, *Equation 1*) and accuracy ( $p < .05$  for both animals, *Equation 3*). Higher absolute coherences  
 117 decreased the RT and increased accuracy, meaning that animals responded faster (*Figure 2b,e*)  
 118 and more accurately (*Figure 2a,d*) on easier trials. Likewise, responses directed towards the large-  
 119 reward target were faster than those directed at the small-reward target (*Figure 2b,e*). Further-  
 120 more, responses were faster and more accurate on trials with shorter presample durations (*Fig-  
 121 ure 2a,b,d,e*).

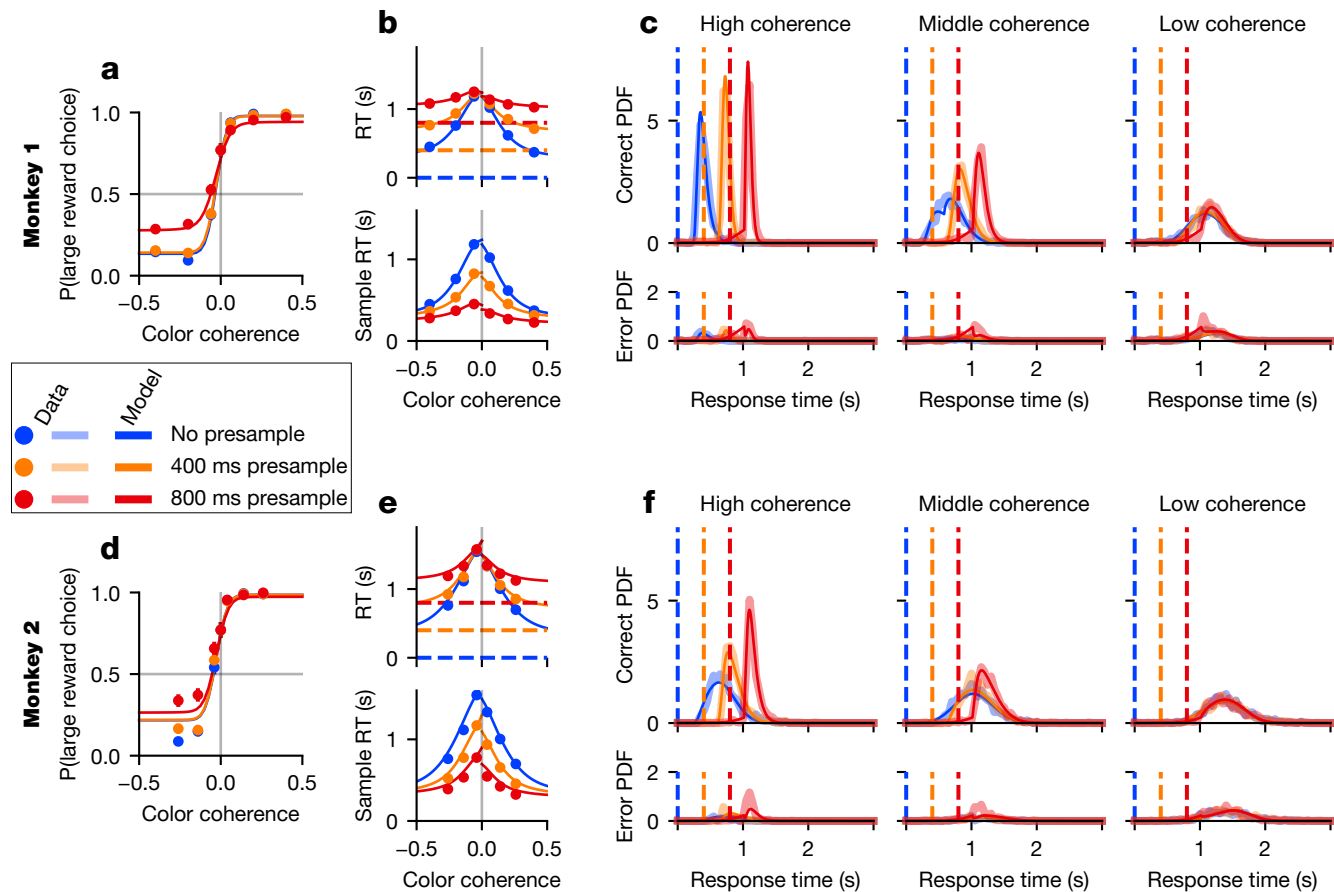
122 We next examined whether these task variables modulated RT and accuracy independently.  
 123 We found that both accuracy ( $p < .05$  for both animals, *Equation 2*) and RT ( $p < .05$  for both ani-  
 124 mals, *Equation 4*) were significantly modulated by the interaction between absolute coherence and  
 125 presample duration. This interaction's effect on RT can be seen in the chronometric function (*Fig-  
 126 ure 2b,e*) by observing that for low-coherence trials, RT aligned to the presample is approximately  
 127 equal regardless of the presample duration. In other words, despite the fact that there is 800ms  
 128 of additional sensory evidence presented to the animal during the 0ms presample trials compared  
 129 to the 800ms presample trials, the RT is similar in both cases. Such similarity in response time is  
 130 reflected in the overlap in the RT distribution in low-coherence trials for all presample durations  
 131 (*Figure 2c,f*). By contrast, for middle- or high-coherence trials, the RT shows a larger effect of co-  
 132 herence.

133 In summary, we find that coherence, presample duration, and large-reward location not only  
 134 modulated RT and accuracy independently, but the interaction between coherence and presample  
 135 duration influences choice and RT. To gain insights into the underlying mechanisms, we tested  
 136 whether a DDM that incorporated these task variables could provide a parsimonious account for  
 137 the observed behaviors.



**Figure 1. Color matching task with asymmetric reward.** The temporal sequence of trial events in the color matching task with asymmetric reward is shown above. Reward cues indicating a large or small reward are inset. Below is a timeline indicating the presence of various task elements on the screen, with presample and sample lines for each condition denoted by the line color.

**Figure 1–video 1. Video of the task.** The task is demonstrated across several coherences and presamples. Element sizes are adjusted for demonstration.



## Drift-diffusion model with reward mechanisms

The DDM is one of the simplest models that formalizes evidence integration for two alternative choices (Luce, 1986; Bogacz et al., 2006; Ratcliff et al., 2016). Recently, studies have shown that reward bias in psychometric functions can be accounted for by two separate mechanisms (Fan et al., 2018; Gesiarz et al., 2019). First, the initial bias shifts the starting position of the decision variable in the direction of the large-reward target (Edwards, 1965; Laming, 1968; Ratcliff, 1985). Second, a time-dependent bias allows a constant evidence signal to be continuously added to the integrator in the direction of the large-reward choice (Mulder et al., 2012; Ratcliff, 1985; Ashby, 1983). These two types of reward biases have been shown to be effective in reproducing choice bias (van Ravenzwaaij et al., 2012; Voss et al., 2004; Diederich and Busemeyer, 2006; Ratcliff and McKoon, 2008).

We thus took this combined reward-biased DDM (Fan et al., 2018; Gesiarz et al., 2019) as a point of departure to gain further insights and understanding of our data. In this extended seven-parameter model (see Methods, Equation 7), the weight on evidence is constant over time, as are the bounds. We fit this model by maximizing likelihood using the method of differential evolution (Storn and Price, 1997).

While the DDM with combined reward bias correctly predicted the large-reward target would be chosen with a higher probability and a shorter RT than the small-reward target, overall this model could not fit other major features of the choice and RT data as measured by the psychometric (Figure 2–Figure Supplement 1a,d) and chronometric functions (Figure 2–Figure Supplement 1b,e) as well as the RT distributions (Figure 2–Figure Supplement 1c,f). First, it failed to reproduce the overlapping RT distributions across different durations of presample for low-coherence samples (Figure 2–Figure Supplement 1c,f). This model, like DDM models in general, predicts that RT should depend on the amount of evidence which has been integrated. However, the overlapping RT distributions indicate a minimal effect of evidence integrated during the early part of the trial compared to evidence later in the trial.

Second, the model failed to produce the characteristic shapes of RT distributions in the data. In both the data and the model, color coherence and sample time jointly influenced the shape of the RT distribution. However, the distribution of the actual RT data became more symmetric as color coherence decreased, whereas the distribution of RT in the model remained skewed similarly in all conditions (Figure 2–Figure Supplement 1c,f top panels).

Finally, the model failed to fit the time course of errors induced by the strong reward bias. During the task, errors during high-coherence stimuli were made almost exclusively because animals tended to choose the large-reward target when the evidence supported the small-reward choice (Figure 2–Figure Supplement 1a,d). Therefore, errors during the high-coherence sample particularly reflect animals' strong bias toward large-reward target and the error rate tended to sharply increase immediately after the sample onset (Figure 2–Figure Supplement 1c,f bottom panels). By contrast, the DDM predicted that fewer errors must follow the onset of the sample, as more evidence supporting correct choice becomes available after sample onset. These results collectively suggest that animals' behavior during the task used in our study cannot be fully explained by a time-invariant evidence integration strategy in conjunction with simple extensions to incorporate reward bias.

## Elements of Generalized DDMs

The above results suggest that relatively simple modifications of the DDM are insufficient in that they cannot accurately reproduce the RT distribution. This is especially true of their ability to model the evolution of reward effects over time under temporally uncertain conditions. Therefore, we extended the DDM framework to create a generalized drift-diffusion model (GDDM) for testing potential mechanisms to account for animals' behavior in our task. GDDMs allow the drift rate and diffusion coefficient to be arbitrary functions of time and of the position of integrated evidence (i.e. decision variable), and integration bounds to be arbitrary functions of time. GDDMs also allow



for a leaky integrator, meaning the model can incorporate both integrative and non-integrative strategies that utilize instantaneous evidence (*Cisek et al., 2009; Thura et al., 2012*) (see Methods, *Equation 9* for the details of the model). Therefore, the GDDM enables fitting of drift-diffusion models that implement complex time-varying mechanisms for evidence sampling and integration as well as reward bias.

### Mechanisms for modulating the time scale of evidence integration

While the simple DDM assumes perfect integration in which the integrated evidence does not decay over time (i.e. “forgetting”), we consider the possibility that animals might integrate evidence at a relatively short time scale so that the integrated evidence primarily reflects the most recent evidence (*Usher and McClelland, 2001; Cisek et al., 2009; Brown and Holmes, 2001; Feng et al., 2009; Brown et al., 2005; Ossmy et al., 2013; Veliz-Cuba et al., 2016*). Therefore, the GDDM included a “leak” parameter that determines the time constant of the decay in decision variable. A leak parameter permits the model to capture both perfect integration, where all evidence is considered in the decision-making process, and non-integration, where only momentary evidence is considered (*Cisek et al., 2009*), thereby generalizing across these distinct strategies.

### Mechanisms of time-varying urgency

To explain animals’ strategic timing of decisions, we hypothesized that animals might dynamically modulate the evidence accumulation process as a function of time during the trial. Within the DDM framework, this was implemented through a time dependence of parameter values, which has been most studied in the context of an “urgency” signal. An urgency signal can be characterized across two dimensions: its type, and its form. Two types of urgency signal include decision bounds which collapse over time (*Ditterich, 2006b; Drugowitsch et al., 2012; Hanks et al., 2014; Murphy et al., 2016*) and a time-dependent increase in the gain of the momentary sensory evidence (*Cisek et al., 2009; Gold and Shadlen, 2001; Ditterich, 2006b*). Here, we consider both of these time-varying urgency signal types as means to modulate the evidence accumulation process.

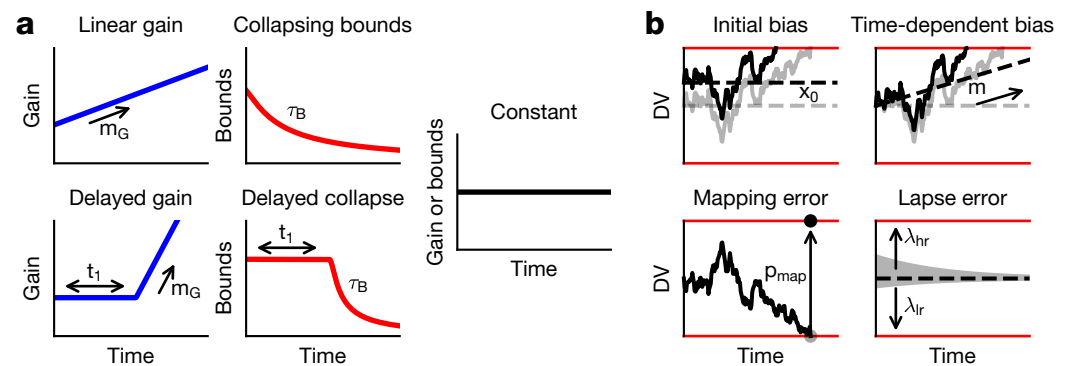
In addition to the urgency signal’s type, we can also consider its form. Consistent with prior work, animals might have begun increasing their urgency at the beginning of the trial. We implemented this as a linearly increasing gain function (“linear gain”, *Figure 3a*) or a gradually decreasing bound (“collapsing bounds”, *Figure 3a*). The limiting cases—where the slope of the gain function is zero or the time constant of the collapsing bound is infinite—correspond to time-independent urgency, or equivalently to the lack of explicit urgency in the simple DDM (“constant”, *Figure 3a*).

In a simple form of a task-specific time-varying urgency signal, animals might have delayed increasing urgency until the uncertainty of evidence onset was sufficiently resolved. Therefore, in addition to urgency signals that begin to ramp at the onset of the sensory stimulus, as is commonly described in the literature, we additionally considered urgency signals which do not begin to ramp until some later point within the trial. We tested both gain modulation and collapsing bounds urgency signals with such a nonlinear time course to allow for an effect of task structure on urgency. The delayed urgency signal introduced a delay before the linear increase in the gain function (“delayed gain”, *Figure 3a*) or exponential collapse of the bounds (“delayed collapse”, *Figure 3a*). This delay parameter would be tuned by the temporal uncertainty of the behavioral task.

### Mechanisms of reward bias

We found that animals’ bias toward the large-reward choice increased as the sample was presented at more predictable time, particularly for higher coherence samples (*Figure 2a,d*). We considered two classes of mechanism, one in which reward directly influences the integration process, and the other in which the reward bias is implemented outside the integration process.

Similar to how the DDM was extended to incorporate reward bias, the effect of reward on the integration process can be modeled by an “initial bias”, which maintains a fixed magnitude throughout the trial, or a “time-dependent bias”, which increases in magnitude throughout the trial



**Figure 3. Generalized DDM components.** Our GDDM included extensions related to timing and reward. (a) We modeled timing using two types of urgency signals. We implemented a gain function and collapsing bounds, each with and without a time delay. Constant gain and constant bounds indicates the absence of an urgency signal. Bounds are shown in red, and gain functions in blue. (b) Reward mechanisms are shown with example decision variable trajectories for each mechanism. Initial bias: the integrator starts biased towards the large-reward target, and the leaky integrator decays back to this starting position instead of to the origin. Time-dependent bias: there is a gradual increase in baseline evidence towards the large-reward target over time. Mapping error: once a decision is reached, the monkey chooses the opposite target on a percentage of trials. Lapse error: there is a higher, exponentially-distributed probability of making an evidence-independent choice to the large-reward target at any given point throughout the trial, contrasted to equal probabilities in the absence of this mechanism.

(Fan et al., 2018; Gesiarz et al., 2019) (Figure 2–Figure Supplement 1; Figure 3b). The initial bias is traditionally implemented as an initial value of the decision variable, but due to leaky integration, this is ineffective because the decision variable will decay from this starting position back to zero. Thus, in addition to setting the initial value of the decision variable, our implementation of initial bias also causes the decision variable to leak towards this initial position. For similar reasons, our implementation of the time-dependent bias is a linear increase in the value to which the decision variable leaks over the course of the trial (see Methods).

We also considered two additional types of reward bias which are outside the integration process. First, the animal may be biased in its categorical choice at the end of evidence accumulation (Erlich et al., 2015; Hanks et al., 2015). Namely, when the decision variable reaches the bound, animals may with some probability mistakenly produce a motor response towards the large-reward choice even if they correctly reached the bound for small-reward choice (“mapping error”, Figure 3b). Second, we also considered that a small number of response may be generated randomly anytime during the integration process (“lapse trials”). The responses on these lapse trials may be biased toward the large-reward choice (Simen et al., 2009; Noorbalooshi et al., 2015), which can be implemented by an asymmetric lapse rate for the large- and small-reward choices (“lapse bias”, Figure 3b) (Yartsev et al., 2018).

### Robust estimation of model parameters

As the models described here have many parameters, methods for parameter estimation must be carefully considered. First, we must consider the metric by which potential models are evaluated. We used a state-of-the-art simulation environment which allows fitting models using maximum likelihood on the full probability distribution. Further details are described in Section Fitting method.

Second, we must protect against overfitting. Before performing any analyses on the data for the task with asymmetric reward, the data were split in half by pseudo-randomly choosing half of the trials for the “validation” set and excluding these from further analysis, analyzing only the disjoint “exploration” data. The validation trials were not fit to the model or otherwise examined until all analyses for the present manuscript were complete. After unmasking the validation trials, no additional analyses were performed on the data. More details are described in Section Cross-



validation procedure.

# **GDDM accounts for strategies**

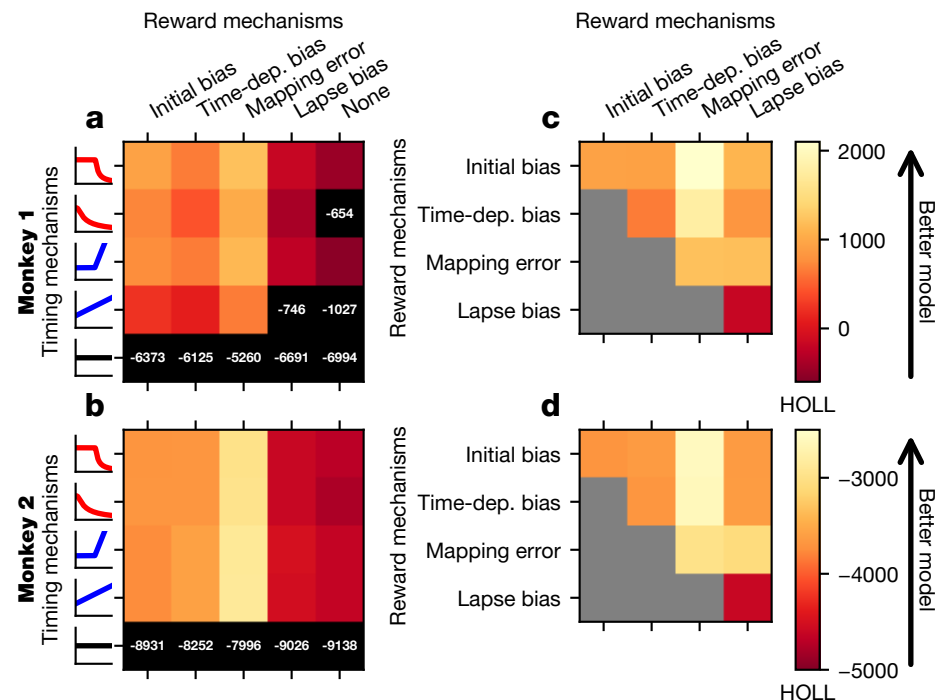
We fit the GDDM to the choice and RT behavior for each animal, and evaluated model fit with held-out log-likelihood (HOLL). Overall, models with time-varying urgency performed remarkably better than ones with constant urgency such as the simple DDM (**Figure 4a,b**). When combined with each of several different reward bias mechanisms, models with delayed urgency best explained the data for monkey 1, while models with non-delayed and delayed urgency performed similarly well for monkey 2. Despite some individual differences, these results suggest that the animal's strategic adaptation to the temporal structure of sensory evidence can be well accounted for by a time-varying urgency.

Regarding reward bias mechanisms, we first compared GDDMs with only one of the four reward bias mechanisms, and found that the models with mapping error performed best. This suggests that asymmetric rewards affected the animal's choice through an integration-independent mechanism (**Figure 4a,b**). Moreover, models with a time-dependent bias did not perform better than those with initial bias, suggesting that the reward bias mechanisms need not depend on time in order to explain the observed effects. We next asked whether these bias mechanisms could be combined to further improve model fit. In conjunction with a delayed collapsing bounds urgency signal, we tested all combinations of two reward bias mechanisms, and found that adding initial bias to the mapping error mechanism consistently improved the model fit for both monkeys more than any other mechanism such as time-dependent bias or lapse bias (**Figure 4c,d**). While initial bias and time-dependent bias both performed well individually and with mapping error, they did not improve the fit of the model when considered together (**Figure 4c,d**). These results suggest that both integration-dependent and integration-independent mechanisms are needed to explain animals' reward bias.

We next examined the role of leak in evidence accumulation in the GDDM. For our best-fit model, the time constant was 140 ms for monkey 1 and 63 ms for monkey 2. We found that given urgency mechanisms, the estimated time scale of integration was consistently short (<150 ms) across all the tested models (**Figure 5**). Model comparison showed that adding a leak parameter to each model improved model fit for all the models except the simple DDM (**Figure 5**), for which the best-fitting leak time constant was often infinity. Notably, the models with non-delayed urgency and the models implementing time-dependent bias mechanism showed particularly large improvement with the addition of the leak parameter (**Figure 5a,b**). This result suggests that the leaky integration might improve the performance of these models by providing a mechanism to disregard early uncertain evidence, a property which can also be represented through a delayed urgency signal. However, the fact that the leak parameter also produced substantial improvement in delayed urgency models suggests that the short time scale of integration itself is an important feature.

The best-fit model captured behavior quite well. It fit the psychometric and chronometric functions and exhibited major features of choice and RT data that cannot be accounted for by simple DDM with integration-based reward mechanisms alone (**Figure 2**). In particular, the best-fit model dramatically improves the quantitative and qualitative fit to the RT distributions, compared to the simpler model (**Figure 2**). This model provided the best fit of all models considered here for Monkey 1, and nearly the best fit for Monkey 2, for which a slightly higher likelihood can be obtained through the use of the delayed gain function as an urgency signal (**Figure 4**).

As mentioned above, the RT distributions in our data during the low-coherence sample overlap with one another for all presample durations, and this could not be captured by the simple DDM with reward bias which uniformly integrates sensory evidence (**Figure 2-Figure Supplement 1**). By contrast, the best GDDM correctly predicts overlapping RT distributions as well as the RT-accuracy relationship for low-coherence samples (**Figure 2**). When evidence is weak, delayed urgency combined with a short time scale of integration implies the decision is most likely to be driven not by



**Figure 4. Fit of potential GDDM models.** Various models were fit to the exploration data from monkey 1 (a,c) or monkey 2 (b,d) and subsequently evaluated with HOLL on the validation data. Color indicates the HOLL of the model, with higher values indicating a better fit. (a,b) Models were constructed by selecting one urgency signal (Figure 3b) and one reward mechanism (Figure 3a). Models with HOLL outside the range of the color bar are shown in black with their corresponding overlaid HOLL value. (c,d) Models were constructed using the delayed collapse urgency signal in conjunction with one or two reward mechanisms (Figure 3a). Models with only one reward mechanism appear on the diagonal. All models used leaky integration.

**Figure 4—Figure supplement 1. Fit of GDDM models with exploration data.** Format similar to Figure 4, except fit is evaluated using BIC on the exploration data instead of HOLL on the validation data.

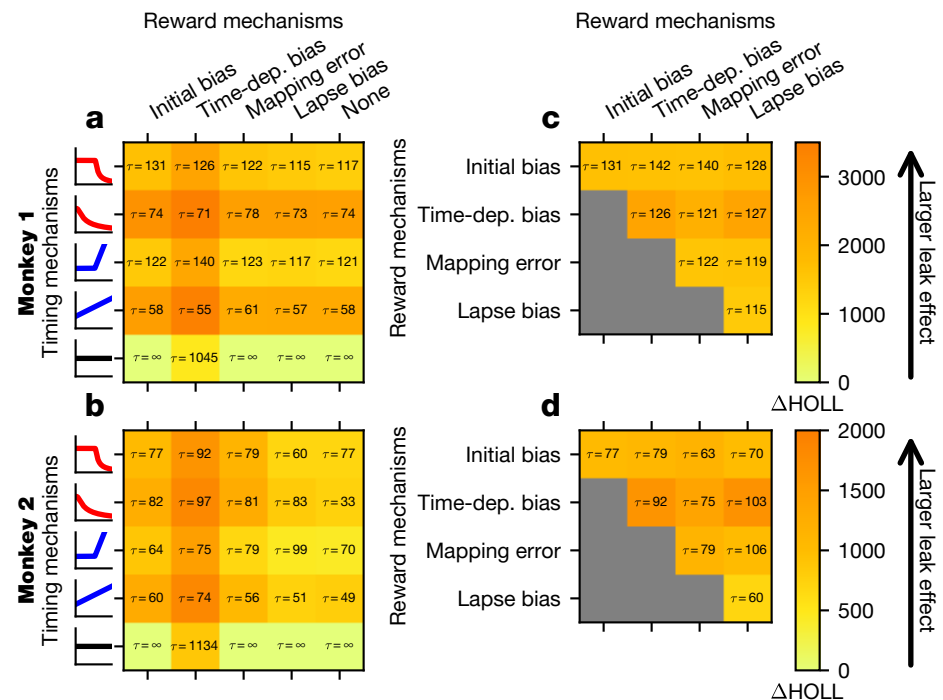
the total evidence, but rather by the later evidence close to the maximum presample duration used (0.8 s), i.e. when the uncertainty about stimulus onset is resolved.

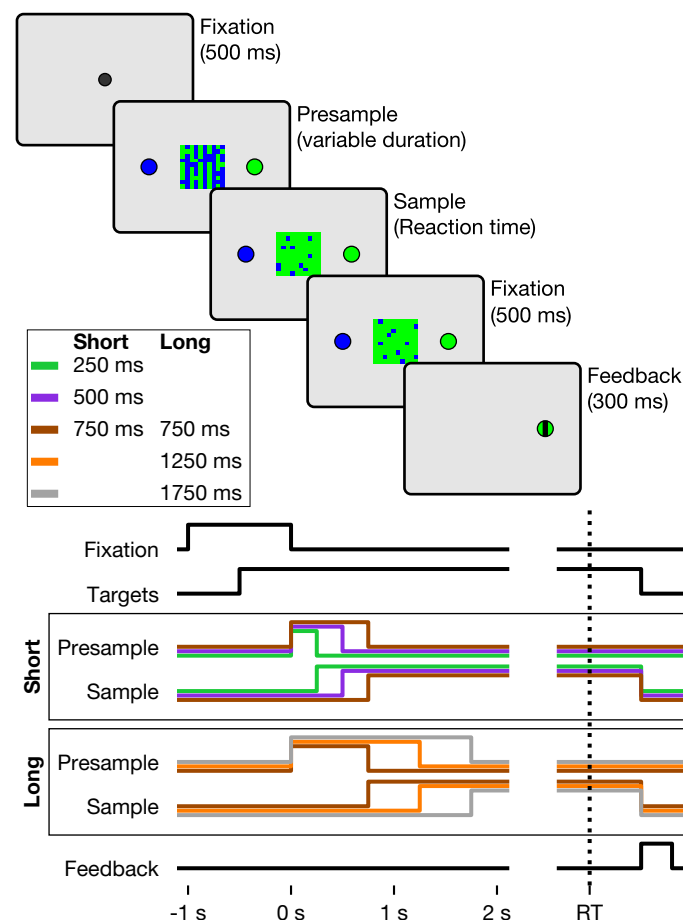
Finally, the best GDDM fit the time course of errors frequently occurring after the onset of the high-coherence sample. This pattern of errors is inconsistent with one of the basic tenets of the DDM that stronger sensory evidence for one alternative makes the subject more likely to choose that alternative. To the contrary, our result shows that in the presence of asymmetric rewards, strong evidence for one alternative can paradoxically make the subject more likely to choose the other alternative. The apparently dynamic reward bias that increases with presample duration is mostly captured by the mapping error mechanism, indicating a failure of mapping the decision variable to motor output correctly.

### Manipulating temporal expectation changes strategy

Thus far, our findings suggest a set of mechanisms in the GDDM to quantitatively capture the behavioral choice and RT patterns. The two temporal features in this model, leaky integration and a delayed urgency signal, hint that the temporal structure of the task may drive their properties. By manipulating the task's structure, we tested whether the urgency signal could be flexibly influenced by temporal context within a single experimental session.

We investigated this question using a variant of the task in the same monkeys (Figure 6). In this new task, the duration of temporal uncertainty was manipulated by varying the set of alternative presample durations between different blocks of trials within a session. In the "short-presample"





**Figure 6. Color matching task with timing blocks.** The temporal sequence of trial events in the color matching task with timing blocks is shown above. Below is a timeline indicating the presence of various task elements on the screen for either the short- or long-presample blocks, with presample and sample lines for each condition denoted by the line color.

blocks, there was a 0.5 s presample in majority of the trials (60%); in the remaining 40% of trials, the sample arrived either earlier (0.25 s) or later (0.75 s) with 20% probability each. Similarly, in the "long-presample" block, 60% of trials had a presample of duration 1.25 s, and 20% each had 0.75 s and 1.75 s (Figure 6). Accordingly, 20% of trials in each block had a 0.75 s presample, which corresponded to the longest presample duration in the short-presample block and the shortest presample duration in the long-presample block. No reward bias was imposed in this modified version of the task, so a correct response to either target elicited an identical reward.

This task was utilized to test whether temporal context can modify the urgency mechanisms of the fitted GDDM. Under a pure DDM framework, the RT distribution for the 0.75 s presample should be the same in both blocks. However, if temporal context influenced behavior, the RTs for the 0.75 s presample in the short-presample block would be on average shorter than those in the long-presample block.

We first performed a linear regression analysis in order to determine the effect of coherence, presample duration, and block on RT and accuracy (see Methods). As expected, we found higher coherences and shorter presample durations both decreased RT ( $p < .05$ , Equation 5) and increased accuracy ( $p < .05$ , Equation 6). More importantly, we also found a significant effect of the block regressor, in that trials presented during the long-presample blocks had significantly longer RT ( $p < .05$ , Equation 5) and higher accuracy ( $p < .05$ , Equation 6) compared to the short-presample

blocks.

To examine more mechanistically how the animals altered their decision-making strategies, we fit the best GDDM chosen for the first task, including a leaky delayed collapsing bound, to the second task. Mechanisms for reward bias were removed given the lack of reward asymmetry in the present task. This model provided an excellent fit to the psychometric and chronometric functions as well as the probability distribution (*Figure 7*).

To investigate which model parameters could best explain the difference between blocks, we allowed individual parameters in this model to differ for the two blocks and assessed the log-likelihood in each case. For comparison, we also fit a 16-parameter model in which all of the parameters were allowed to vary between blocks, and an 8-parameter model which fixed all parameters to be the same for both blocks.

We examined the difference in HOLL when each parameter was fit separately for each block compared to shared between the blocks, as well as when all parameters or no parameters are shared. We found that the 16-parameter model fit the data better than models in which only a single parameter was allowed to vary between the short and long blocks. Nevertheless, changes in the urgency signal alone were sufficient to explain most of this improvement in model fit (*Figure 8*). For example, in *Figure 7*, the model which allows the “collapse delay”  $t_1$  parameter to vary is shown. This demonstrates that the timing of the urgency signal mediates a critical aspect of the monkey’s change in strategy between different temporal contexts.

## Urgency timing can implement a speed-accuracy tradeoff

The above results showed that context-dependent changes in the timing of the urgency signal can account for the changes in behavior related to temporal uncertainty. In theory, the animal could increase accuracy by simply delaying the onset of the increase in urgency signal for the longest possible duration (1750 ms) on every trial, which would minimize the effect of integrating noise. However, this would come at the cost of long RT. Conversely, the animal could reduce RT by beginning to increase urgency immediately at the beginning of the trial, but at the cost of lower accuracy. Therefore, the onset of the urgency signal can mediate a speed-accuracy tradeoff, with longer onset delays favoring accuracy and shorter onset delays favoring speed.

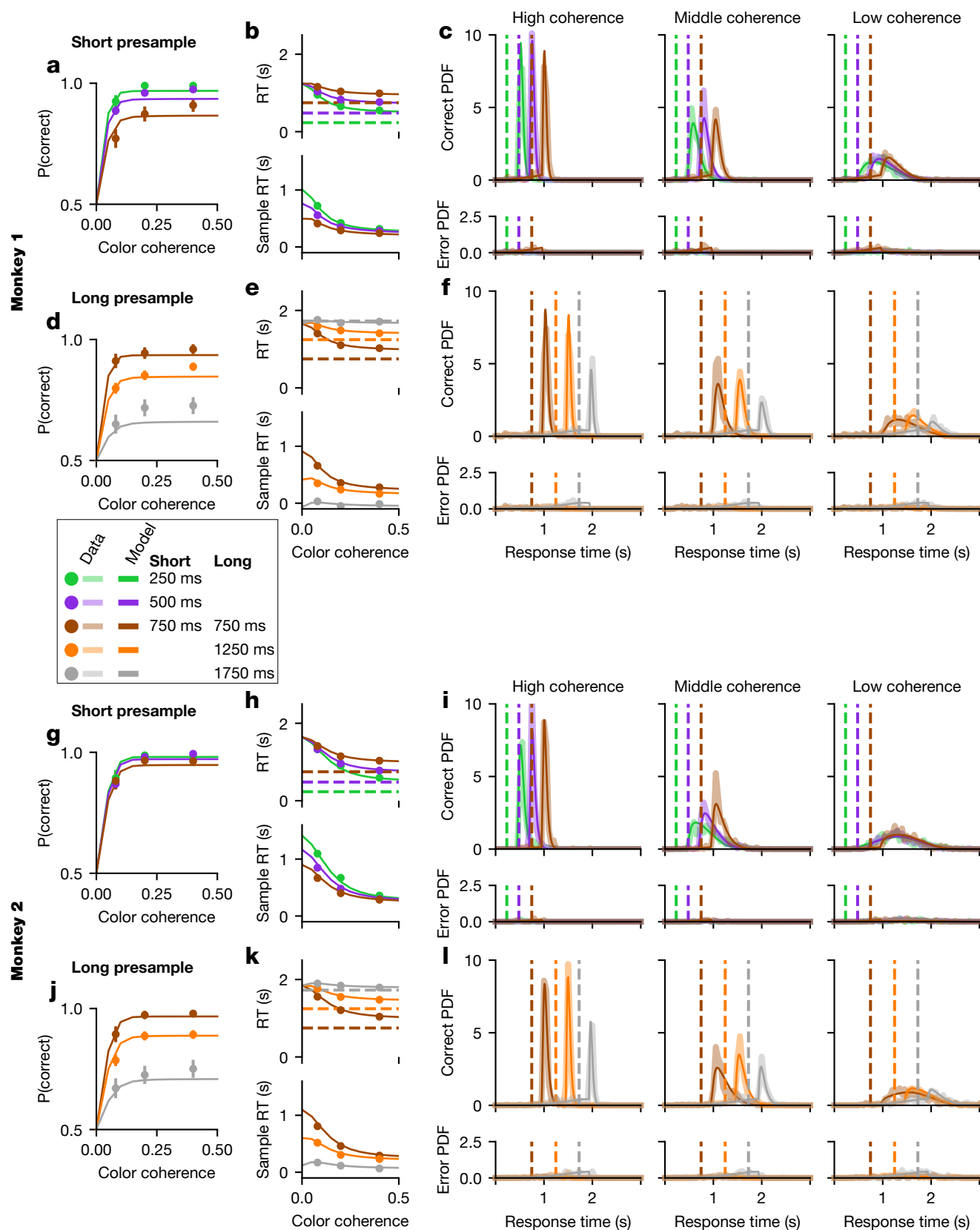
To examine this quantitatively, we systematically varied the “collapse delay”  $t_1$  (*Figure 3b*) from 0 ms to 1000 ms, separately for long- and short-presample blocks, to observe how this parameter modulated the speed-accuracy tradeoff in the GDDM, with all other parameters shared (*Figure 9*). This analysis confirmed that changes to the urgency signal are able to control the speed-accuracy tradeoff. Furthermore, the speed-accuracy tradeoff is strikingly similar for both monkeys, despite the differences in estimated parameter values. This demonstrates that strategic modulation of speed-accuracy tradeoff can be accomplished using changes to only the urgency signal, which suggests that urgency may be critical for timing-related decision strategies.

## Discussion

In this study, we found that perceptual decision-making is driven by temporal changes in evidence quality via a dynamic urgency signal. We also showed that reward bias in animals’ choice can be accounted for by integration-dependent and integration-independent mechanisms.

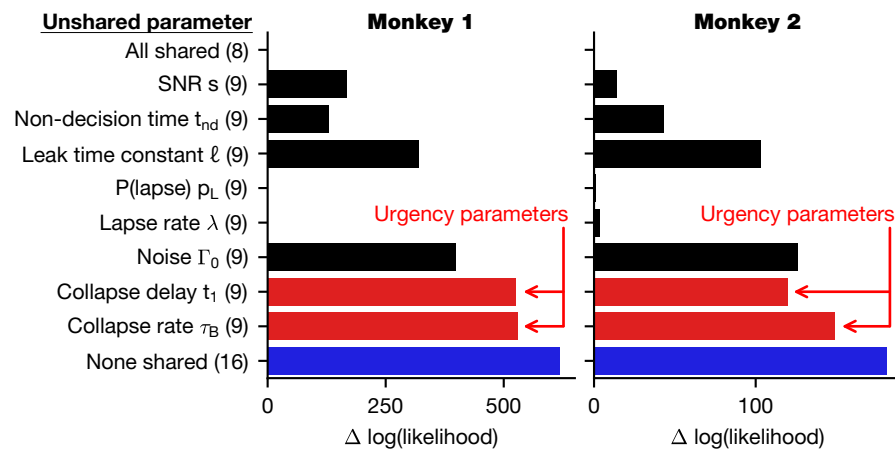
## Temporal uncertainty and decision-making

An urgency to commit to a decision has long been hypothesized to modulate the speed and accuracy of perceptual decision-making (*Reddi and Carpenter, 2000*). In computational models, urgency has been implemented either by a collapsing bound or gain modulation of sensory evidence (*Ditterich, 2006a,b; Forstmann et al., 2008, 2010; Ratcliff and McKoon, 2008; Churchland et al., 2008; Cisek et al., 2009; Drugowitsch et al., 2012; Hanks et al., 2014; Murphy et al., 2016*). In models with collapsing bounds, an evidence-independent, time-varying signal is combined with the accumulators with fixed weight on evidence, while the gain function multiplicatively modulates the evidence

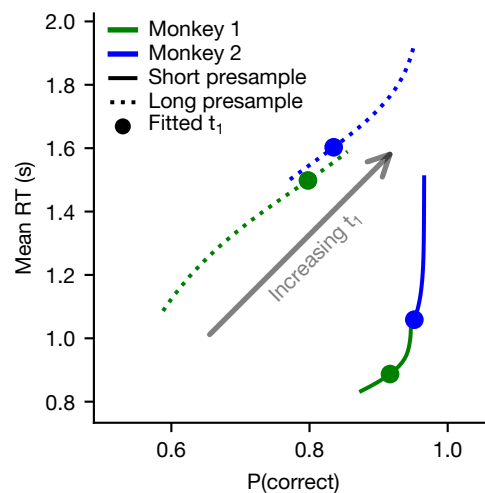


**Figure 7. Model fit to timing dataset.** Psychometric functions (a,d,g,j), chronometric functions (b,e,h,k), and RT histograms (c,f,i,l) are shown for short-presample (a,b,c,g,h,i) and long-presample (d,e,f,j,k,l) blocks for monkey 1 (a,b,c,d,e,f) and monkey 2 (g,h,i,j,k,l). Data are represented by dots in the behavioral functions and thick translucent lines in the probability distributions. “Sample RT” indicates the RT minus the presample duration. Overlaid as a thin solid line is a GDDM including delayed collapsing bounds and a leaky integrator with a separate “collapse delay”  $t_1$  parameter for short- and long-presample blocks. Error bars for 95% CI are hidden beneath the markers.





**Figure 8. Urgency signal captures changes between presample blocks.** An eight-parameter GDDM with a delayed collapse urgency signal and leaky integration was fit to data from the color matching task with presample blocks. The block type (either long- or short-presample) was incorporated into the model by allowing one parameter at a time to vary between blocks, forming a 9 parameter model, and then re-fitting all parameters of the new model to the data. For comparison, we also fit the 16 parameter model where we allowed all parameters to differ between blocks (blue). The improvement in log likelihood for each model is shown, as well as the total number of parameters of the model in parentheses. Monkey 1 is shown on the left, and monkey 2 on the right. Parameters related to the urgency signal are highlighted in red.



**Figure 9. Urgency signal mediates a speed-accuracy tradeoff.** A GDDM with a delayed collapse urgency signal and leaky integration was fit to each monkey's RT data in the color matching task with presample blocks. Using these parameters, the "collapse delay  $t_1$ " parameter was varied systematically across the linearly-spaced values from 0 to 1, representing the hypothetical case in which the subject could control this parameter while leaving the other parameters fixed. For each value, the mean RT and probability of a correct response were plotted for the long-presample (dotted lines) and short-presample (solid lines) blocks. The actual parameters are shown as filled dots.

with time-varying weight. However, the nature of time-varying mechanisms has been controversial (*Hawkins et al., 2015a,b; Voskuilen et al., 2016*). During the task used in our study, we varied the uncertainty about evidence onset over time to probe the time-varying mechanisms. Using an array of computational models, we showed that an urgency signal provides a flexible mechanism by which temporal information can be incorporated into the decision.

Recently, several studies have found, using dynamically changing sensory evidence, that subjects can adjust the gain or weight on the evidence across time during the trial, depending on the temporal statistics of evidence (*Cheadle et al., 2014; Levi et al., 2018*). In these previous studies, however, subjects were not allowed to freely choose the timing of their response, so it was not possible to determine how the subject's strategy reflected a speed-accuracy trade-off. In the present study, we adopted an RT paradigm and explicitly showed that animals voluntarily adjusted the timing of their decision in order to reduce uncertainty about evidence through task-specific time-varying urgency. While we did not explore the potential optimality of this strategy, this presents an interesting opportunity for future study (*Drugowitsch et al., 2012, 2014*).

Time-varying stimuli were also used to test the time scale of evidence integration (*Bronfman et al., 2016*). Whereas the traditional DDM assumed a perfect integrator, recent studies have raised possibility that evidence integration may be leaky, as the accuracy of the decision does not necessarily increase with integration time (*Usher and McClelland, 2001; Zariwala et al., 2013; Tsetsos et al., 2015; Farashahi et al., 2018*). Moreover, when evidence dynamically changes during the course of a trial, models with leaky integration or instantaneous evidence fit the data well, as shown in the present study (*Cisek et al., 2009; Thura et al., 2012; Thura and Cisek, 2014; Bronfman et al., 2016*). In addition to leaky integration, bounded integration has also been proposed as a mechanism to account for the apparent failure to use the full stream of stimulus information (*Kiani et al., 2008*).

Our use of a dynamic urgency signal can be seen as a generalization of the idea that integration must begin at a single point in time (*Teichert et al., 2016*). Recently, integration onset was examined by *Devine et al. (2019)*, who tested human subjects on a perceptual decision-making paradigm with multiple presample durations, a single sample coherence level, and no reward bias. Based on behavioral and EEG data, they proposed that under temporal uncertainty, integration begins approximately at the time of the earliest sample onset. However, their analyses did not fit a model of decision-making processes to test against alternative mechanisms. By contrast, we find strong evidence that our subjects used the temporal statistics of the task to modulate the evidence accumulation process. Applying our model comparison framework in other task paradigms can test the generality of these strategies.

A monotonically increasing gain function provides a straightforward mechanism to weight late evidence more than early evidence. The collapsing bounds mechanism does not by itself weight late evidence more than early evidence, although the interplay between time-varying urgency and leaky integration may provide a mechanism for temporal weighting of late evidence over early evidence. For example, in fixed-duration paradigms, a leaky integrator can effectively weight late evidence more than early evidence (*Levi et al., 2018; Lam et al., 2017*). We found that leaky integration improved fit in all models but played a larger role for simpler urgency signals than for the best-fitting delayed urgency signals, suggesting that leaky integration may capture a related component of behavior as the delay in urgency increase. While our results imply that both leaky integration and a delayed urgency signal are necessary, the ability of a time-varying urgency signal to reduce the relative benefit of adding leaky integration to the model demonstrates the flexibility of the urgency signal.

## Asymmetric rewards and decision-making

Asymmetry in reward or prior probability of the correct target can induce a bias to targets with a higher expected value (*Voss et al., 2004; Lauwereyns et al., 2002; Diederich and Busemeyer, 2006; Ratcliff and McKoon, 2008; Mulder et al., 2012; Feng et al., 2009; Rorie et al., 2010; Teichert and Fer-*

449 *vera, 2010*). Several studies have indicated that reward bias in perceptual decision-making could  
450 be well accounted for by an integration-dependent mechanism in which reward is modeled either  
451 a constant evidence (i.e. initial bias) or momentary evidence (i.e. time-dependent bias) (*Feng et al.,*  
452 *2009; Rorie et al., 2010; Fan et al., 2018; Gesiarz et al., 2019*). In our task, even a combination of  
453 these two reward mechanisms could not fit our data in the absence of an urgency signal or leaky  
454 integration. Likewise, recent studies have reported that under time pressure, reward bias might  
455 be produced by integration-independent mechanisms such as a quick guess that races with the  
456 process of evidence integration (*Simen et al., 2009; Noorbalooshi et al., 2015; Afacan-Seref et al.,*  
457 *2018*). However, this possibility was not consistent with our data either. This discrepancy might be  
458 due to the fact that previous studies have seldom tested the possible role of urgency in modulating  
459 the balance between evidence and bias when determining choice. Although reward bias increas-  
460 ing with time following the onset of an informative sensory stimulus could imply a time-varying  
461 bias mechanism, our results showed that such a time-varying bias is not necessary to explain our  
462 data. Overall, our results suggest that temporal and reward structures play an important role in  
463 perceptual decision-making. In addition, models within the GDDM framework can effectively de-  
464 scribe the contribution of temporal and reward related factors during decision-making and play  
465 an important role in investigating the cognitive mechanisms of decision-making.

466 As investigators begin to explore more complicated behavioral tasks and computational mod-  
467 els, it is important to use methods of model-fitting which are specialized for complex, high-dimensional  
468 models. In this study, we fit each model to the entire RT distributions of correct and error choices  
469 using maximum likelihood, and were able to provide qualitatively good fits to both the RT distribu-  
470 tion and the psychometric and chronometric functions. By contrast, much previous work has fit  
471 models directly to the psychometric and chronometric functions, producing excellent fits to these  
472 reduced behavioral functions with a potentially poor fit to the RT distributions. Indeed, we found  
473 that evaluating and providing model fits to the entire probability distribution are highly informative  
474 for differentiating performance of different models and mechanisms. This methodology is imper-  
475 ative for simulating and fitting complex models. Our methodological innovations will therefore  
476 allow for increased complexity in experimental design to test a broad range of hypotheses about  
477 the mechanisms of decision-making.

## 478 **Methods**

### 479 **Animal preparation**

480 Two male rhesus macaque monkeys (Q and P, identified as monkey 1 and 2 respectively: body  
481 weight, 10.5–11.0 kg) were used. The animal was seated in a primate chair with its head fixed  
482 and eye position was monitored with a high-speed eye tracker (ET49, Thomas Recording, Giessen,  
483 Germany). All procedures used in this study were approved by the Institutional Animal Care and  
484 Use Committee at Yale University, and conformed to the Public Health Service Policy on Human  
485 Care and Use of Laboratory Animals and Guide for the Care and Use of Laboratory Animals.

### 486 **Behavioral tasks**

487 Animals were trained to perform a color matching task, in which the onset time of the sample  
488 was systematically manipulated by varying the duration of non-informative stimulus (presample)  
489 that preceded the sample. Two versions of the task were used. In the color matching task with  
490 asymmetric reward (*Figure 1*), either large or small magnitude of reward was randomly assigned  
491 for identifying the correct color of the sample at a given trial. Presample duration was randomly  
492 selected from three discrete values at each trial. In the color matching task with timing blocks  
493 (*Figure 6*), two sets of presample intervals were used and presented in alternating blocks of trials  
494 (i.e. short- and long-presample blocks). In each block, presample duration was randomly selected  
495 from three discrete values. The magnitude of reward was the same for all correct choices. Animals  
496 were first trained for the task with timing blocks and then for the task with asymmetric reward.

## Color matching task with asymmetric reward

Each trial began when the animal fixated a central square (white,  $0.6^\circ \times 0.6^\circ$  in visual angle) (**Figure 1**). After the 0.3 s fixation period, cues indicating magnitudes of reward available from two alternative choices were presented at the potential locations of saccade targets. Two sets of reward cues were used: in addition to those shown in **Figure 1**, for some sessions, a circle was used for the large-reward target and a diamond for the small-reward target. Two nested squares in thin gray lines (or a circle) and a single square in thicker white lines (or a  $45^\circ$ -angled square) indicated larger and smaller reward, respectively. The ratio of reward magnitude was 2:3 and 1:4 for monkey 1 and 2, respectively. Following the reward-cue period (0.4 s), a green disk was presented inside one of the two reward cues and a blue disk appeared within the other, serving as targets for choice saccade. After 0.4 s of target period, a checkered square with equal numbers of green and blue pixels ( $20 \times 20$ ;  $1.63^\circ \times 1.63^\circ$  in visual angle), or presample, was presented replacing the central fixation target. In addition, green and blue pixels were dynamically redistributed to random positions at a rate of 20 Hz. After a variable interval (0, 0.4, 0.8 s), the presample was replaced by the sample, which was identical to the presample except that it contained green and blue pixels in different proportions. Presample duration was randomly sampled from the three intervals with equal probability. The animal was allowed to shift its gaze towards one of the peripheral target at any time after the onset of the sample and was rewarded only when it correctly chose the target with the same color as the majority of the pixels in the sample. Choices made during presample were punished by an immediate end of the trial followed by a 2 s timeout. Three levels of color coherence were used for the sample (53%, 60%, 70% for monkey 1; 52%, 57%, 63% of pixels for monkey 2), each with equal probability. In about 5% of the trials, the presample never changed to a sample, meaning the coherence of each color was 50% and there was no correct choice. On these trials, a reward was delivered with 50% probability regardless of the monkey's choice. After 0.3 s of fixation on the chosen target, feedback was provided for 0.3 s with a horizontal or vertical bar superimposed on the chosen target indicating erroneous or correct choice, respectively. Reward was delivered for correct choices. Color coherence, location of saccade target with correct color, presample duration, and location of large-reward cue were pseudo-randomly selected for each trial. Monkey 1 performed 30868 trials across 50 sessions, and monkey 2 performed 23695 trials across 29 sessions.

## Color matching task with timing blocks

The sequence of trial events was the same as in the color matching task with asymmetry (**Figure 6**), except that fixation and saccade targets were presented simultaneously without the reward cue, and the durations of the presample varied between blocks. Presample duration for the majority of the trials (60%) in the short- and long-presample block was 0.5 s and 1.25 s, respectively. In the remaining trials, the presample was terminated with equal probability (20%) either earlier (0.25 s and 0.75 s for short- and long-presample blocks, respectively) or later (0.75 s and 1.75 s for short- and long-presample blocks, respectively) than the most frequent presample durations. Choices made during the presample were penalized by subsequent 5 s timeout. Three levels of color coherence were used for the sample (54%, 60% and 70% for both monkey 1 and 2), each with equal probability. Short- and long-presample blocks alternated within a given session and the order of the two types of block was randomized across sessions. Monkey 1 performed 12,773 trials across 10 sessions, and monkey 2 performed 10,777 trials across 10 sessions.

## Fitting method

As we fit models with many parameters, we must be careful about the methods used for parameter estimation. One common approach in the field is to fit the parameters of the DDM based only on the probability of a correct response and the mean RT for correct trials. This uses only a very small subset of the data, using only two summary statistics, and is thus insufficient for fitting complex models. A second approach which is commonly used when fitting complex models is to simulate

individual trajectories of the decision variable, and then perform tests on the quadrants of the simulated values to determine whether there is a match. While this approach utilizes more data than the previously discussed method, it still uses only summary statistics of the data. More critically, this simulation process is very slow, and hence does not permit efficiently fitting the model to data. These two approaches are principally used to overcome limitations in simulating and fitting these models, namely, that a full RT distribution is not always available. Our approach simulates the stochastic differential equations by numerically solving the Fokker-Planck equations so that we may use likelihood-based methods on the entire RT distribution for estimating parameters. This allows all data to be utilized through a robust statistical framework.

RT distributions simulated directly from the Fokker-Planck equation often use the forward Euler method, an easy-to-implement method for solving stochastic differential equations. The forward Euler method mandates the use of very small time steps and a coarse decision variable discretization to maintain numerical stability. In practice, this means that simulations which achieve a reasonable margin of error are prohibitively time-consuming for fitting parameters.

To circumvent this, we instead solve the Fokker-Planck equation using the Crank-Nicolson method (for the fixed-bound conditions) or the backward Euler method (for the collapsing bound conditions). These methods do not require small time steps in order to achieve low margin of error. As a result, we are able to fit parameters using the results of these numerical solutions. Simulations were performed using the PyDDM package<sup>1</sup> using a timestep of 5 ms and decision variable discretization of 0.005. Correctness of the implementation was verified using specialized techniques (Shinn, 2019).

Fitting is performed by maximizing the log-likelihood function. For low-dimensional models, parameters may be optimized using a variety of local search functions which tend to give similar results. However, for high-dimensional models, these methods are often unable to minimize the function, instead finding a local minimum or failing to converge. Because we were simultaneously fitting up to 16 parameters, using an appropriate optimization routine was critical. We used differential evolution, a heuristic search method as implemented in Scipy (Jones et al., 2001). Differential evolution is a global search method, meaning it is intended to avoid local minima in the search space, and has been shown to exhibit consistent results in practice (Storn and Price, 1997). This gave consistent parameter estimates and similar log-likelihoods for our GDDMs.

## Cross-validation procedure

Once obtaining a fit, there are many different processes which may be used for evaluating the fit of a model to protect against overfitting. There are two types of overfitting which are important to consider for complex models such as ours with many parameters. First, overfitting may occur because the model mechanism is complicated and thus the parameters of the model may fit to noise within the data. Second, overfitting may be the result of the modeler testing many potential model mechanisms and choosing the best mechanism. Simple models with few parameters may only be concerned with the first of these, but models with complicated mechanisms must guard against both types of overfitting.

Many papers simply evaluate the best fit and report these values. This does not protect against either overfitting the parameters or overfitting the model mechanism. A more sophisticated approach is to utilize a measure such as AIC or BIC, which takes the model complexity into account when comparing two competing models. However, these metrics judge model complexity exclusively by the number of parameters, and this has been shown to be a poor metric for evaluating model complexity (Piantadosi, 2018; Myung et al., 2000). Additionally, this only helps with overfitting the parameter values and does not prevent overfitting of the model mechanism. A better approach is to utilize n-fold cross validation, which will properly penalize complex models over simple models by ensuring that the parameters are not overfit. However, this still leaves open

<sup>1</sup><https://github.com/mwshinn/PyDDM>

the possibility for overfitting the model mechanism. For complex models, the best way to prevent overfitting is to use held-out data which is not analyzed or otherwise examined in any way until the final model has been constructed and the conclusions of the model have been determined.

Therefore, in order to ensure robustness and protect against both types of overfitting, we split the dataset into two halves, an “exploration” dataset and a “validation” dataset. All analyses and model fitting were performed only on the exploration data using Bayesian information criterion (BIC) for comparing models. The validation trials were not fit to the model or otherwise examined until all analyses for the present manuscript were complete. After unmasking the validation trials, no additional analyses were performed, no additional results were added to the manuscript, and no existing results were removed. These analyses are shown as supplementary figures (*Figure 2–Figure Supplement 2*, *Figure 4–Figure Supplement 1*, and *Figure 5–Figure Supplement 1*). For a more robust analysis, we evaluate models on the validation dataset using the held-out log-likelihood (HOLL), i.e. the likelihood under the held-out (validation) data when parameters were fit using the exploration dataset (*Figure 2*, *Figure 4*, *Figure 5*). All quantitative and qualitative results found using BIC were unchanged when using HOLL with the validation dataset.

We did not perform this validation procedure on the task with timing blocks, as sample size was limited. Nevertheless, all models were developed on data from the task with asymmetric reward before being applied to data from the task with timing blocks.

## Behavioral analyses

In analyzing animals’ behavior, we measured response time (RT) relative to the presample onset. However, we occasionally refer to the animal’s response time relative to the onset of the sample (“Sample RT”).

A linear model was fit to examine the dependence of RT on task parameters:

$$\text{Response time} = \beta_0 + \beta_1 R + \beta_2 D + \beta_3 |C| \quad (1)$$

Here,  $R$  is reward magnitude, where 1 indicates that the large-reward target was the correct response and 0 that the small-reward target was correct;  $D$  is presample duration in units of seconds; and  $|C|$  is unsigned coherence, where 1 corresponds to a solid color patch and 0 to an equal ratio of each color. We define RT relative to the onset of the presample, rather than the sample, given that the transition from presample to sample was not explicitly cued. Coherence takes values greater than or equal to 0. All notation is described in *Table 1*.

To examine second-order interaction terms, a second model was fit:

$$\text{Response time} = \beta_0 + \beta_1 R + \beta_2 D + \beta_3 |C| + \beta_4 RD + \beta_5 R|C| + \beta_6 |C|D \quad (2)$$

Similarly, a binomial generalized linear model was fit to understand how the task parameters influenced the monkey’s decision to choose the correct target:

$$\text{logit}(\mathbb{P}(\text{correct choice})) = \beta_0 + \beta_1 R + \beta_2 D + \beta_3 |C| \quad (3)$$

where variables are as defined above. Likewise, a second model was fit which contained second-order interaction terms:

$$\text{logit}(\mathbb{P}(\text{correct choice})) = \beta_0 + \beta_1 R + \beta_2 D + \beta_3 |C| + \beta_4 RD + \beta_5 R|C| + \beta_6 |C|D \quad (4)$$

Linear model results were similar in both the exploration and validation datasets. Trials with zero coherence were excluded.

Likewise, RT was examined for the color match task with timing blocks using the linear model:

$$\text{Response time} = \beta_0 + \beta_1 D + \beta_2 |C| + \beta_3 B \quad (5)$$



**Table 1. Summary of notation.** All mathematical notation used in equations and models are listed below.

**Task variables**

$D$	Presample duration in ms
$C$	Coherence strength from -1 to +1, $C > 0$ indicates large-reward target correct
$R$	1 = large-reward target correct, 0 = small-reward target correct
$B$	1 = long-presample block, 0 = short-presample block

**Parameters**

$s$	Signal to noise ratio ( $s > 0$ )
$\ell$	Leak magnitude, 0 = perfect integration, $\infty$ = no integration ( $\ell \geq 0$ )
$t_{nd}$	Non-decision time, ( $t_{nd} \geq 0$ )
$\mu$	Drift rate (DDM only), ( $\mu \geq 0$ )
$\sigma$	Noise (DDM only), ( $\sigma > 0$ )
$p_L$	The probability of a lapse, ( $p_L \geq 0$ )
$\lambda$	Decay rate of lapse trials (symmetric) ( $\lambda > 0$ )
$t_1$	Onset time of increase in urgency ( $t_1 \geq 0$ )
$\Gamma_0$	Urgency at time $t = 0$ ( $\Gamma_0 > 0$ )
$m_G$	Slope of linear ramp in gain function ( $m_G \in \mathbb{R}$ )
$\tau_B$	Decay rate of exponential decrease in bounds ( $\tau_B > 0$ )
$x_0$	Starting position of decision variable ( $x_0 \in \mathbb{R}$ )
$m$	Increase in leak baseline over time ( $m \in \mathbb{R}$ )
$p_{map}$	Probability of a mapping error ( $p \geq 0$ )
$\lambda_{hr}$	Decay of lapse probability towards the large-reward target (replaces $\lambda$ ) ( $\lambda_{hr} \geq 0$ )
$\lambda_{lr}$	The decay rate of lapse trials towards the small-reward target (replaces $\lambda$ ) ( $\lambda_{lr} \geq 0$ )

**Misc**

$x$	Decision variable
$\Gamma(t)$	Gain function
$\Theta(t)$	Decision bound
$I$	Indicator function
$W$	Wiener process
$G$	1 = correct trial, 0 = error trial
$\beta_i$	Regression coefficients
$\tau$	The leak time constant, defined as $\frac{1}{\ell}$

**Table 2. Model parameters.** Parameters for each displayed model. Rew: Color matching task with asymmetric reward, Time: Color matching task with timing blocks.

Task:	Rew	Rew	Time	Time	Time	Time	Rew	Rew
Model:	GDDM	GDDM	GDDM	GDDM	GDDM	GDDM	DDM	DDM
Monkey:	1	2	1	1	2	2	1	2
Block:	—	—	Short	Long	Short	Long	—	—
$s$	9.32	6.52	11.25	11.25	12.11	12.11		
$\Gamma_0$	1.03	1.16	1.62	1.62	1.40	1.40		
$t_{nd}$	0.22	0.21	0.20	0.20	0.22	0.22	0.27	0.21
$\ell$	7.14	15.82	22.67	22.67	9.81	9.81		
$\tau_B$	1.19	0.70	0.35	0.35	0.47	0.47		
$t_1$	0.36	0.09	0.00	0.21	0.20	0.73		
$x_0$	0.09	0.05					-0.16	-0.04
$p_{map}$	0.12	0.21						
$p_L$	0.04	0.03	0.03	0.03	0.04	0.04	0.03	0.10
$\lambda$	1.02	0.12	0.47	0.47	1.47	1.47	0.48	0.95
$\mu$							2.90	6.27
$\sigma$							0.77	0.76
$m_e$							0.60	0.43

where  $B$  is the block, coded as 0 for the short-presample block and 1 for long-presample, and other terms are as specified above. Accuracy was likewise analyzed with the binomial generalized linear model:

$$\text{logit}(\mathbb{P}(\text{correct choice})) = \beta_0 + \beta_1 D + \beta_2 |C| + \beta_3 B \quad (6)$$

### DDM with reward mechanisms

The DDM is governed by a decision variable  $x$ . The evolution of  $x$  in our reward-biased DDM model, similar to that of *Fan et al. (2018)* and *Gesiarz et al. (2019)*, is described by:

$$dx = I_{t>D} \mu C dt + m_e dt + \sigma dW \quad (7)$$

where the following parameters were fit to data:

- $\mu$  - signal strength
- $\sigma$  - noise level
- $m_e$  - momentary sensory evidence bias

Additional terms in the model are:

- $W$  - A Wiener process with standard deviation 1
- $C$  - The color coherence
- $I_{t>D}$  - An indicator function for the presample on the current trial.  $I_{t>D} = 0$  if we are in the presample, and  $I_{t>D} = 1$  otherwise.

We specify that  $x$  has initial condition given by parameter  $x = x_0$  at  $t = 0$  seconds. We use the convention that positive values of  $x$  indicate choices towards the large-reward target, and thus positive coherence indicates motion towards the large-reward target and negative coherence away from it. When  $|\int_0^t dx| \geq 1$ , a decision to choose left or right is made based on the sign of  $x$ . Since the parameter  $\sigma$  is fit to data, integration bounds are fixed at 1 to prevent over-parameterization. The RTs generated by this model were shifted post-simulation by the non-decision time  $t_{nd}$ .

Since we fit based on maximum likelihood, rare trials in which the choice appears unrelated to the stimulus (lapses) may have a large impact on the fitted parameters. Thus, for the purposes

of model fitting, we fit a mixture model between the drift-diffusion process and an exponentially-distributed lapse rate with rate parameter  $\lambda$ . This is given by:

$$f(t) = \frac{1}{2} \lambda e^{-\lambda t} \quad (8)$$

where there is a  $p_L$  probability of any given trial being a lapse trial. While we fit both of these parameters to data, results were unchanged when we fixed these parameters at constant values (data not shown).

Thus, overall, this model contained seven parameters:  $\mu$ ,  $\sigma$ ,  $x_0$ ,  $m_e$ ,  $t_{nd}$ ,  $p_L$ , and  $\lambda$ .

## Generalized DDM

Like the DDM, the GDDM is governed by the evolution of a decision variable  $x$ . The instantaneous value of  $x$  is described by:

$$dx = -\ell (x - x_0 + m t) dt + I_{t>D} \Gamma(t) C s dt + \Gamma(t) dW \quad (9)$$

where the parameters which could potentially be fit by simulations are:

- $\ell$  - The leak parameter, constrained to  $\ell \geq 0$ . Its inverse is the leak time constant.
- $x_0$  - The initial position of the integrator, and the location to which the leaky integrator decays. This was constrained to be  $x_0 \geq 0$ . By default, it was fixed to  $x_0 = 0$ .
- $m$  - The change in the leaky integrator baseline over time. By default this was fixed at  $m = 0$ .
- $\Gamma(t)$  - The sensory gain, a function of time. By default,  $\Gamma(t) = \Gamma_0$ .
- $s$  - The signal-to-noise ratio, constrained to  $s \geq 0$ .
- $\Theta(t)$  - The decision bound. When  $|\int_0^t dx| \geq \Theta(t)$ , a decision to choose left or right is made based on the sign of  $x$ . Since noise is allowed to vary, integration bounds are subject to the constraint  $\Theta(0) = 1$  to prevent overfitting. In the absence of collapsing bounds, this fixes the bounds to be equal to  $\pm 1$ , the default.

Additional terms in this equation are defined as:

- $W$  - White noise, i.e. a Wiener process.
- $C$  - Coherence. We use the convention that positive values of  $x$  indicate choices towards the large-reward target, and thus positive coherence indicates motion towards the large-reward target and negative coherence away from it.
- $I_{t>D}$  - An indicator function for the presample on the current trial.  $I_{t>D} = 0$  if we are in the presample, and  $I_{t>D} = 1$  otherwise.

As above, we fit a mixture model between the drift-diffusion process and an exponentially-distributed lapse rate with rate  $\lambda$ , with a  $p_L$  probability of any given trial being a lapse trial. This is given by:

$$f(t) = \frac{1}{2} \lambda e^{-\lambda t} \quad (10)$$

Overall, in the simplest possible case where all parameters are set to their defaults (thereby indicating that no reward or timing effects are included in the model), the model includes six parameters:  $\lambda$ ,  $p_L$ ,  $\Gamma_0$ ,  $\ell$ ,  $s$ ,  $t_{nd}$ . Reward and timing mechanisms added to the model increased the number of parameters which were fit.

## Reward mechanisms

Within the GDDM framework, we designed reward mechanisms for the model according to our hypothesized cognitive mechanisms, shown in **Figure 3**.

## Initial bias

The total evidence necessary to reach a decision is lower for the large-reward target than for the small-reward target. This is equivalent to changing the starting position of integration. Because our model deals with leaky integration, we furthermore impose that the leak decays to this position instead of to zero. This adds one parameter to the model: the magnitude of the baseline shift  $x_0$ .

## Time-dependent bias

The total evidence necessary to reach a decision is initially the same for the large- and small-reward targets, but a bias towards the large-reward target over the small-reward target develops linearly over time. Without leak, this is equivalent to a constant bias on the drift rate, i.e. a continuous integration of a constant. Because our model deals with leaky integration, such a bias would instead change the point to which the leaky integrator decays. Thus, we model this bias by temporally modulating the position to which the leaky integrator decays. This adds one parameter to the model: the slope of the increase in large-reward bias  $m$ .

## Mapping error

After integrating to the bound, with some probability, small-reward choices are mapped to a response toward the large-reward target. Equivalently, some percentage of small-reward choices are assigned to be large-reward instead. This mechanism was first described in discrete evidence paradigms (Erlich *et al.*, 2015; Hanks *et al.*, 2015). This mechanism adds one parameter to the model: the probability of making a mapping error on any given trial  $p_{map}$ .

Mapping error was implemented as post-simulation RT histogram modifications. For a simulated probability density function  $f(x|R = r, C = c, G = g, D = p)$ , where  $R$  is large vs. small reward,  $C$  is coherence,  $G$  is correct or error trial, and  $D$  is the presample duration, we compute the final density  $f'(x)$  using  $f(x)$ . For the mapping error, this was calculated as:

$$f'(x|R = \text{large}, C = c, G = \text{corr}, D = d) = f(x|R = \text{large}, C = c, G = \text{corr}, D = d) + p_{map} f(x|R = \text{large}, C = c, G = \text{err}, D = d) \quad (11)$$

$$f'(x|R = \text{large}, C = c, G = \text{err}, D = d) = (1 - p_{map}) f(x|R = \text{large}, C = c, G = \text{err}, D = d) \quad (12)$$

$$f'(x|R = \text{small}, C = c, G = \text{corr}, D = d) = (1 - p_{map}) f(x|R = \text{small}, C = c, G = \text{corr}, D = d) \quad (13)$$

$$f'(x|R = \text{small}, C = c, G = \text{err}, D = d) = f(x|R = \text{small}, C = c, G = \text{err}, D = d) + p_{map} f(x|R = \text{small}, C = c, G = \text{corr}, D = d) \quad (14)$$

## Lapse rate bias

The lapse rate, assumed to be exponentially distributed, is allowed to be higher in the direction of the large-reward target compared to the small-reward target. This is given for large- and small-reward targets respectively by:

$$f_{hr}(t) = \lambda_{hr} e^{-(\lambda_{hr} + \lambda_{lr})t} \quad (15)$$

$$f_{lr}(t) = \lambda_{lr} e^{-(\lambda_{hr} + \lambda_{lr})t} \quad (16)$$

This mechanism has three parameters: the lapse rates for both large- and small-reward targets  $\lambda_{hr}$  and  $\lambda_{lr}$ , and the probability of any given trial being a lapsed trial  $p_L$ . Models which include this mechanism do not include unbiased lapse rate, as it is redundant. Since the unbiased lapse rate uses two parameters, this mechanism adds one net parameter to the model.

## Timing mechanisms

Similarly, we designed timing mechanisms to capture the ideas of urgency signals. Two types of urgency signal have been previously described in the literature: a “gain function” which scales evidence and noise uniformly throughout the course of the trial, and “collapsing bounds” which cause the decision bounds to become less stringent as the trial progresses.

## Collapsing bounds

Collapsing bounds can be used to implement an urgency signal of various forms. While much debate has focused on the correct form of the bounds (Hawkins et al., 2015a; Malhotra et al., 2017), we use an exponentially collapsing bound for simplicity. This mechanism adds one parameter to the model: the time constant of the exponential collapse  $\tau_B$ . It is implemented as:

$$\Theta(t) = e^{-t/\tau_B} \quad (17)$$

Note that this form satisfies the imposed boundary condition  $\Theta(0) = 1$ .

## Delayed collapsing bounds

This implements an exponentially decreasing bound as an urgency signal similarly to collapsing bounds, except the bounds do not begin to collapse until late in the trial. This mechanism adds two parameters to the model: the time before the bounds begin to collapse  $t_1$ , and the rate of collapse  $\tau_B$ . This is implemented as:

$$\Theta(t) = \begin{cases} e^{-(t-t_1)/\tau_B}, & t \geq t_1 \\ 1, & \text{otherwise} \end{cases} \quad (18)$$

These also satisfies the boundary condition  $\Theta(0) = 1$  for all  $t_1 \geq 0$ .

## Linear gain

A gain function can also be used to implement an urgency signal. It scales both the evidence and the noise simultaneously in order to preserve signal-to-noise ratio. This mechanism adds one parameter to the model: the rate at which gain increases  $m_G$ . It is implemented as:

$$\Gamma(t) = \Gamma_0 + m_G t \quad (19)$$

## Delayed gain

The delayed gain function implements an urgency signal similarly to linear gain but it does not begin to increase the gain until part way into the trial. This mechanism adds two parameters to the model: the time at which the value begins to ramp  $t_1$ , and the slope of the ramp  $m_G$ . It is implemented as:

$$\Gamma(t) = \begin{cases} \Gamma_0 + m_G (t - t_1), & t \geq t_1 \\ \Gamma_0, & \text{otherwise} \end{cases} \quad (20)$$

## Simulations without leak

Simulations without leak were performed by setting  $\ell = 0$ .

## Acknowledgments

We thank Norman H. Lam for assistance with model fitting.

## References

- Afacan-Seref, K., Steinemann, N. A., Blangero, A., and Kelly, S. P. (2018). Dynamic interplay of value and sensory information in high-speed decision making. *Current Biology*, 28(5):795–802.e6.
- Ashby, F. (1983). A biased random walk model for two choice reaction times. *Journal of Mathematical Psychology*, 27(3):277–297.
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9):1214–1221.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and 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):700–765.

- 762 Bronfman, Z. Z., Brezis, N., and Usher, M. (2016). Non-monotonic temporal-weighting indicates a dynamically  
763 modulated evidence-integration mechanism. *PLOS Computational Biology*, 12(2):e1004667.
- 764 Brown, E., Gao, J., Holmes, P., Bogacz, R., Gilzenrat, M., and Cohen, J. D. (2005). Simple neural networks that  
765 optimize decisions. *International Journal of Bifurcation and Chaos*, 15(03):803–826.
- 766 Brown, E. and Holmes, P. (2001). Modeling a simple choice task: stochastic dynamics of mutually inhibitory  
767 neural groups. *Stochastics and Dynamics*, 01(02):159–191.
- 768 Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Castañón, S. H., and Summerfield, C. (2014). Adap-  
769 tive gain control during human perceptual choice. *Neuron*, 81(6):1429–1441.
- 770 Churchland, A. K., Kiani, R., and Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature*  
771 *Neuroscience*, 11(6):693–702.
- 772 Cisek, P., Puskas, G. A., and El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model.  
773 *Journal of Neuroscience*, 29(37):11560–11571.
- 774 Devine, C. A., Gaffney, C., Loughnane, G., Kelly, S. P., and O’Connell, R. G. (2019). The role of premature evidence  
775 accumulation in making difficult perceptual decisions under temporal uncertainty.
- 776 Diederich, A. and Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual  
777 discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psy-*  
778 *chophysics*, 68(2):194–207.
- 779 Ding, L. and Gold, J. I. (2012). Separate, causal roles of the caudate in saccadic choice and execution in a  
780 perceptual decision task. *Neuron*, 75(5):865–874.
- 781 Ditterich, J. (2006a). Evidence for time-variant decision making. *European Journal of Neuroscience*, 24(12):3628–  
782 3641.
- 783 Ditterich, J. (2006b). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural*  
784 *Networks*, 19(8):981–1012.
- 785 Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., and Pouget, A. (2014). Optimal multisensory  
786 decision-making in a reaction-time task. *eLife*, 3.
- 787 Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., and Pouget, A. (2012). The cost of accumu-  
788 lating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11):3612–3628.
- 789 Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times,  
790 and human information processing. *Journal of Mathematical Psychology*, 2(2):312–329.
- 791 Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., and Brody, C. D. (2015). Distinct effects of prefrontal and  
792 parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife*, 4.
- 793 Fan, Y., Gold, J. I., and Ding, L. (2018). Ongoing, rational calibration of reward-driven perceptual biases. *eLife*, 7.
- 794 Farashahi, S., Donahue, C. H., Khorsand, P., Seo, H., Lee, D., and Soltani, A. (2017). Metaplasticity as a neural  
795 substrate for adaptive learning and choice under uncertainty. *Neuron*, 94(2):401–414.e6.
- 796 Farashahi, S., Ting, C.-C., Kao, C.-H., Wu, S.-W., and Soltani, A. (2018). Dynamic combination of sensory and  
797 reward information under time pressure. *PLOS Computational Biology*, 14(3):e1006070.
- 798 Feng, S., Holmes, P., Rorie, A., and Newsome, W. T. (2009). Can monkeys choose optimally when faced with  
799 noisy stimuli and unequal rewards? *PLoS Computational Biology*, 5(2):e1000284.
- 800 Forstmann, B. U., Anwander, A., Schafer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., Bogacz, R., and Turner,  
801 R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making.  
802 *Proceedings of the National Academy of Sciences*, 107(36):15916–15920.
- 803 Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., and Wagenmakers,  
804 E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National*  
805 *Academy of Sciences*, 105(45):17538–17542.
- 806 Gao, J., Tortell, R., and McClelland, J. L. (2011). Dynamic integration of reward and stimulus information in  
807 perceptual decision-making. *PLoS ONE*, 6(3):e16749.



- 808 Gesiarz, F., Cahill, D., and Sharot, T. (2019). Evidence accumulation is biased by motivation: A computational  
809 account. *PLOS Computational Biology*, 15(6):e1007089.
- 810 Gold, J. I. and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends*  
811 *in Cognitive Sciences*, 5(1):10–16.
- 812 Grosjean, M., Rosenbaum, D. A., and Elsinger, C. (2001). Timing and reaction time. *Journal of Experimental*  
813 *Psychology: General*, 130(2):256–272.
- 814 Hanks, T., Kiani, R., and Shadlen, M. N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque  
815 area LIP. *eLife*, 3.
- 816 Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., and Brody, C. D. (2015). Distinct relationships  
817 of parietal and prefrontal cortices to evidence accumulation. *Nature*, 520(7546):220–223.
- 818 Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., and Shadlen, M. N. (2011). Elapsed decision time affects the  
819 weighting of prior probability in a perceptual decision task. *Journal of Neuroscience*, 31(17):6339–6352.
- 820 Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., and Brown, S. D. (2015a). Revisiting the evi-  
821 dence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*,  
822 35(6):2476–2484.
- 823 Hawkins, G. E., Wagenmakers, E.-J., Ratcliff, R., and Brown, S. D. (2015b). Discriminating evidence accumulation  
824 from urgency signals in speeded decision making. *Journal of Neurophysiology*, 114(1):40–47.
- 825 Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python.
- 826 Kiani, R., Hanks, T. D., and Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions  
827 even when viewing duration is dictated by the environment. *Journal of Neuroscience*, 28(12):3017–3029.
- 828 Krajbich, I. and Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual  
829 fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, 108(33):13852–  
830 13857.
- 831 Lam, N. H., Borduqui, T., Hallak, J., Roque, A. C., Anticevic, A., Krystal, J. H., Wang, X.-J., and Murray, J. D. (2017).  
832 Effects of altered excitation-inhibition balance on decision making in a cortical circuit model.
- 833 Laming, D. (1979). Autocorrelation of choice-reaction times. *Acta Psychologica*, 43(5):381–412.
- 834 Laming, D. R. J. (1968). *Information theory of choice-reaction times*. Academic Press.
- 835 Lauwereyns, J., Watanabe, K., Coe, B., and Hikosaka, O. (2002). A neural correlate of response bias in monkey  
836 caudate nucleus. *Nature*, 418(6896):413–417.
- 837 Levi, A. J., Yates, J. L., Huk, A. C., and Katz, L. N. (2018). Strategic and dynamic temporal weighting for perceptual  
838 decisions in humans and macaques. *eneuro*, 5(5):ENEURO.0169–18.2018.
- 839 Los, S. A. (2010). Foreperiod and sequential effects: theory and data. *Attention and time*, 289(302):24.
- 840 Luce, R. D. (1986). *Response Times*. OXFORD UNIV PR.
- 841 Malhotra, G., Leslie, D. S., Ludwig, C. J. H., and Bogacz, R. (2017). Time-varying decision boundaries: insights  
842 from optimality analysis. *Psychonomic Bulletin & Review*, 25(3):971–996.
- 843 Mormann, M. M., Malmaud, J., Huth, A., Koch, C., and Rangel, A. (2010). The drift diffusion model can account  
844 for the accuracy and reaction time of value-based choices under high and low time pressure. *SSRN Electronic*  
845 *Journal*.
- 846 Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., and Forstmann, B. U. (2012). Bias in the brain: A  
847 diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, 32(7):2335–2343.
- 848 Murphy, P. R., Boonstra, E., and Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent  
849 urgency during perceptual choice in humans. *Nature Communications*, 7:13526.
- 850 Myung, I. J., Balasubramanian, V., and Pitt, M. A. (2000). Counting probability distributions: Differential geome-  
851 try and model selection. *Proceedings of the National Academy of Sciences*, 97(21):11170–11175.

- 852 Nagano-Saito, A., Cisek, P., Perna, A. S., Shirdel, F. Z., Benkelfat, C., Leyton, M., and Dagher, A. (2012). From  
853 anticipation to action, the role of dopamine in perceptual decision making: an fMRI-tyrosine depletion study.  
854 *Journal of Neurophysiology*, 108(2):501–512.
- 855 Newsome, W. T., Britten, K. H., and Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*,  
856 341(6237):52–54.
- 857 Noorbalooshi, S., Sharon, D., and McClelland, J. L. (2015). Payoff information biases a fast guess process in  
858 perceptual decision making under deadline pressure: Evidence from behavior, evoked potentials, and quan-  
859 titative model comparison. *Journal of Neuroscience*, 35(31):10989–11011.
- 860 Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., and Donner, T. H. (2013). The timescale of perceptual  
861 evidence integration can be adapted to the environment. *Current Biology*, 23(11):981–986.
- 862 Palmer, J., Huk, A. C., and Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of  
863 a perceptual decision. *Journal of Vision*, 5(5):1.
- 864 Piantadosi, S. T. (2018). One parameter is always enough. *AIP Advances*, 8(9):095118.
- 865 Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2):59–108.
- 866 Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses.  
867 *Psychological review*, 92:212–225.
- 868 Ratcliff, R., Cherian, A., and Segraves, M. (2003). A comparison of macaque behavior and superior colliculus  
869 neuronal activity to predictions from models of two-choice decisions. *Journal of Neurophysiology*, 90(3):1392–  
870 1407.
- 871 Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks.  
872 *Neural Computation*, 20(4):873–922.
- 873 Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion decision model: Current issues and  
874 history. *Trends in Cognitive Sciences*, 20(4):260–281.
- 875 Reddi, B. A. J. and Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*,  
876 3(8):827–830.
- 877 Resulaj, A., Kiani, R., Wolpert, D. M., and Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*,  
878 461(7261):263–266.
- 879 Roitman, J. D. and Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined  
880 visual discrimination reaction time task. *The Journal of Neuroscience*, 22(21):9475–9489.
- 881 Rorie, A. E., Gao, J., McClelland, J. L., and Newsome, W. T. (2010). Integration of sensory and reward information  
882 during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS ONE*,  
883 5(2):e9308.
- 884 Rosenbaum, D. A. and Collyer, C. E. (1998). *Timing of behavior: Neural, psychological, and computational perspec-*  
885 *tives*. MIT Press.
- 886 Shinn, M. (2019). Refinement type contracts for verification of scientific investigative software. In *Proceedings*  
887 *of the 11th Working Conference on Verified Software: Theories, Tools, and Experiments*.
- 888 Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., and Cohen, J. D. (2009). Reward rate optimization in two-  
889 alternative decision making: Empirical tests of theoretical predictions. *Journal of Experimental Psychology:*  
890 *Human Perception and Performance*, 35(6):1865–1897.
- 891 Storn, R. and Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization  
892 over continuous spaces. *Journal of Global Optimization*, 11(4):341–359.
- 893 Teichert, T. and Ferrera, V. P. (2010). Suboptimal integration of reward magnitude and prior reward likelihood  
894 in categorical decisions by monkeys. *Frontiers in Neuroscience*, 4.
- 895 Teichert, T., Grinband, J., and Ferrera, V. (2016). The importance of decision onset. *Journal of Neurophysiology*,  
896 115(2):643–661.
- 897 Thura, D., Beauregard-Racine, J., Fradet, C.-W., and Cisek, P. (2012). Decision making by urgency gating: theory  
898 and experimental support. *Journal of Neurophysiology*, 108(11):2912–2930.

- 899 Thura, D. and Cisek, P. (2014). Deliberation and commitment in the premotor and primary motor cortex during  
900 dynamic decision making. *Neuron*, 81(6):1401–1416.
- 901 Tsetsos, K., Pfeiffer, T., Jentgens, P., and Donner, T. H. (2015). Action planning and the timescale of evidence  
902 accumulation. *PLOS ONE*, 10(6):e0129473.
- 903 Usher, M. and McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator  
904 model. *Psychological Review*, 108(3):550–592.
- 905 van Ravenzwaaij, D., Mulder, M. J., Tuerlinckx, F., and Wagenmakers, E.-J. (2012). Do the dynamics of prior  
906 information depend on task context? an analysis of optimal performance and an empirical test. *Frontiers in*  
907 *psychology*, 3:132.
- 908 Veliz-Cuba, A., Kilpatrick, Z. P., and Josić, K. (2016). Stochastic models of evidence accumulation in changing  
909 environments. *SIAM Review*, 58(2):264–289.
- 910 Voskuilen, C., Ratcliff, R., and Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the  
911 diffusion model. *Journal of Mathematical Psychology*, 73:59–79.
- 912 Voss, A., Rothermund, K., and Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical  
913 validation. *Memory & Cognition*, 32(7):1206–1220.
- 914 Woodrow, H. (1914). The measurement of attention. *The Psychological Monographs*, 17(5):i–158.
- 915 Yartsev, M. M., Hanks, T. D., Yoon, A. M., and Brody, C. D. (2018). Causal contribution and dynamical encoding  
916 in the striatum during evidence accumulation. *eLife*, 7.
- 917 Zariwala, H. A., Kepecs, A., Uchida, N., Hirokawa, J., and Mainen, Z. F. (2013). The limits of deliberation in a  
918 perceptual decision task. *Neuron*, 78(2):339–351.

