

Reviewed Preprint

v1 • December 19, 2025

Not revised

Reviewed Preprint

v2 • August 26, 2026

Revised by authors

✉ For correspondence:

amirhosein.hadianrasanan@unibas.ch

Competing interests:

No competing interests declared

Funding:

See [page 31](#)

Reviewing editor:

Redmond G O'Connell, Trinity College Dublin, Ireland

© 2025, Hadian Rasanan et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Non-decision time-informed collapsing threshold diffusion model: A joint modeling framework with identifiable time-dependent parameters

Amir Hosein Hadian Rasanan¹✉, Lukas Schumacher¹, Michael D Nunez², Gabriel Weindel^{3,4}, Jörg Rieskamp¹

¹Department of Psychology, University of Basel, Basel, Switzerland • ²Psychological Methods, University of Amsterdam, Amsterdam, Netherlands • ³Department of Experimental Psychology, Utrecht University, Utrecht, Netherlands •

⁴Institute of Psychology, University of Lausanne, Lausanne, Switzerland

eLife Assessment

This **valuable** study examines whether neurally-derived non-decision time estimates can guide decision models to more reliable parameter estimates. **Convincing** evidence for their value is obtained from state-of-the-art model-fitting methods applied to two empirical datasets. However, the premise relies on very strong assumptions regarding the degree to which the neural metrics directly index non-decision time and comparisons across models that appear to differ in more ways than just the non-decision time constraint.

<https://doi.org/10.7554/eLife.109730.2.sa4>

Abstract

Over the past sixty years, evidence accumulation models have emerged as a dominant framework for explaining the neural and behavioral aspects of the process underlying decision making. These models have also been widely used as a measurement instrument to assess individual differences in latent cognitive constructs underlying decision making. A central assumption of most of these models is that decision makers accumulate noisy evidence until a fixed decision threshold is reached. However, both behavioral and neuroscientific findings, along with theoretical considerations related to optimality, have suggested that the decision threshold varies over time. Although time-dependent threshold models often provide a better account of empirical data, a major challenge associated with these models is the unreliable estimation of their parameters. This limitation has led researchers to emphasize model-fitting comparisons rather than interpreting parameter values or accounting for individual differences in the dynamics of the decision threshold. In this work, we address the reliability issue of parameter estimation in time-dependent threshold diffusion models by proposing a joint modeling approach that links non-decision time to external observations. Parameter recovery simulations demonstrate that informing the diffusion model with trial-level noisy measurements of non-decision time substantially improves the reliability of parameter estimation for time-dependent threshold diffusion models. Additionally, we reanalyzed the experimental data from two perceptual decision-making tasks to illustrate the feasibility of the proposed modeling approach. Non-decision time measurements were extracted from electroencephalography (EEG) recordings using the hidden multivariate pattern method. The cognitive modeling results revealed that, in addition to the reliable parameter estimation, constraining non-decision time improves the absolute fit to behavioral data.

Introduction

For decades, understanding how people make rapid decisions has been a central focus in cognitive science. Studying quick decisions allows researchers to investigate the properties of latent cognitive processes underlying decision making and to develop process-level theories of human choice behavior. Among the various formal cognitive models proposed to explain decision making, evidence accumulation models (EAMs; Ratcliff, 1978; Ratcliff and McKoon, 2008; Stone, 1960; Laming, 1968) represent the most prominent class of cognitive computational models (Forstmann et al., 2016). A key advantage of EAMs is their ability to account for both choice and response time (RT) simultaneously (Forstmann et al., 2016; Ratcliff et al., 2016). The most widely used variant of EAMs is the well-known diffusion decision model (DDM; Ratcliff, 1978; Ratcliff and Rouder, 1998; Ratcliff and McKoon, 2008), originally proposed to explain behavior in two-alternative choice tasks. The DDM assumes that the decision maker accumulates noisy evidence (with the mean evidence accumulation rate governed by a constant *drift rate*) until the relative evidence for one option reaches a fixed decision threshold. This model has made substantial contributions to our understanding of the neural and cognitive mechanisms underlying decision making (e.g., Forstmann et al., 2016; Gold and Shadlen, 2007; Ratcliff et al., 2016). Moreover, it has been widely employed as a cognitive psychometric instrument for data analysis and for describing latent cognitive processes across various domains, including perceptual decisions (e.g., Dutilh and Rieskamp, 2016; Forstmann et al., 2011; Evans and Brown, 2017), risky decisions (e.g., Olschewski et al., 2025; Bhui, 2019; Zhao et al., 2020), value-based decisions (e.g., Krajbich et al., 2010; Fontanesi et al., 2019; Khodadadi et al., 2017; Gluth et al., 2020), numerical cognition (e.g., Ratcliff and McKoon, 2018; Ratcliff, 2022), intelligence (e.g., Schmiedek et al., 2007; Schubert and Frischkorn, 2020), and aging (e.g., Starns and Ratcliff, 2010; Ratcliff, 2008; Ratcliff et al., 2010; von Krause et al., 2022), among others. Additionally, this model has been employed to investigate the information processing patterns in clinical populations (e.g., Nejati et al., 2022; Ging-Jehli et al., 2021; Pirrone et al., 2017; Pedersen et al., 2017; Karalunas et al., 2012). The broad application of the DDM is attributed to the interpretability of its parameters, which has been validated in several studies (e.g., Voss et al., 2004; Arnold et al., 2015; Lerche and Voss, 2019), as well as their reliability in parameter estimation (Lerche and Voss, 2017; Yap et al., 2012).

Although the DDM has become a standard model for describing the latent cognitive constructs of decision making, alternative variants — such as the urgency gating model (Cisek et al., 2009; Thura et al., 2012; Trueblood et al., 2021) and collapsing threshold diffusion models (CT-DDMs; Drugowitsch et al., 2012; Tajima et al., 2016) — better align with certain behavioral and neuroscientific findings (Ratcliff and Frank, 2012; Ging-Jehli et al., 2025). In particular, the DDM assumes that the decision threshold remains fixed over time, implying that the decision maker requires a constant amount of evidence to make a decision, regardless of the time spent on deliberating. However, several neuroscience studies have shown that human and non-human primates tend to become more urgent as time progresses (e.g., Ditterich, 2006; Thura et al., 2012; Cisek et al., 2009; Murphy et al., 2016; Steinemann et al., 2018; Gluth et al., 2012, 2013; Ging-Jehli et al., 2025; Grogan et al., 2025). These findings support the idea of the collapsing threshold or urgency gating models, in which the required amount of evidence for making a decision decreases over time, thereby increasing the urgency of decision making. (Although the urgency gating model and collapsing threshold diffusion model are theoretically distinct, Smith and Ratcliff, 2022 demonstrated that they can be transformed into one another. Therefore, we primarily focus on collapsing threshold models for the remainder of this paper; however, we will explain how the current work can also be related to urgency gating models in more detail in the general discussion.)

CT-DDMs also predict several behavioral effects that fixed-threshold diffusion models (FT-DDMs) fail to capture. First, CT-DDMs are more flexible in accounting for RT distributions with less asymmetry. Specifically, while FT-DDMs typically predict strongly right-skewed RT distributions (Ratcliff and Smith, 2004), CT-DDMs can account for Gaussian-like RT distributions (Evans and Hawkins, 2019; Hawkins et al., 2015). Several studies have observed such symmetric,

Gaussian-like RT distributions under specific conditions (e.g., Roitman and Shadlen, 2002; Ditterich, 2006; Hawkins et al., 2019; Evans and Hawkins, 2019; O’Connell et al., 2012). A second key prediction of CT-DDMs is that slower responses tend to exhibit greater variability in accuracy (in inferential decisions) or in choice proportion (in preferential decisions) — an effect supported by several empirical studies (e.g., Olschewski et al., 2025; Murphy et al., 2016; Steinemann et al., 2018). In other words, these studies found a higher error rate for trials with longer response times. In contrast, the FT-DDM predicts a constant error rate across both fast and slow responses (it is worth noting that including trial-to-trial variability in drift rate enables the FT-DDM to predict slow errors [Ratcliff and Rouder, 1998; Voskuilen et al., 2016]). Beyond these behavioral effects, some empirical studies have shown that experimental manipulations such as introducing a decision deadline or emphasizing urgency encourage individuals to adopt a collapsing threshold strategy (Hawkins et al., 2015; Evans et al., 2020a; Evans and Hawkins, 2019; Katsimpokis et al., 2020). Furthermore, quantitative model comparison showed broad empirical support for the idea that people adjust their decision threshold during a single decision even without time-pressure manipulation (Olschewski et al., 2025; Palestro et al., 2018; Bhui, 2019; Khodadadi et al., 2017; Ging-Jehli et al., 2025; Ratcliff and Frank, 2012) (despite all the aforementioned evidence supporting CT-DDMs, some computational modeling studies have not found substantial improvements in model fit when incorporating a time-dependent threshold compared to a fixed threshold [e.g., Voskuilen et al., 2016; Smith and Ratcliff, 2022]).

On top of the supporting evidence for CT-DDMs from empirical studies, there is also a strong theoretical foundation for these models. CT-DDMs offer a normative, optimal policy for evidence accumulation in several scenarios. The FT-DDM maximizes accuracy for a fixed mean RT — or, equivalently, minimizes mean RT for a fixed accuracy level — in environments with homogeneous choice difficulty and no external biases (Bogacz et al., 2006; Moran, 2015). However, this information processing strategy is not optimal in more complicated settings. For instance, in environments where the difficulty of the choice problem is unknown or varies across trials (Drugowitsch et al., 2012; Tajima et al., 2016; Fudenberg et al., 2018), or where a stochastic deadline governs decision timing (Frazier and Yu, 2007), or where external biases (e.g., prior information about the correct response) are present (Moran, 2015), CT-DDMs provide the optimal evidence accumulation policy.

Although there is both theoretical and empirical support for CT-DDMs, a major limitation that restricts their application as measurement models for exploring individual differences is that some CT-DDM parameters cannot be reliably estimated. Several systematic evaluations using various parameter estimation methods have reported poor parameter recovery for CT-DDMs (Evans et al., 2020b; Murrow and Holmes, 2024a; Fengler et al., 2021), especially when the threshold declines nonlinearly. In other words, these studies found that the actual generating parameters of CT-DDMs are not recoverable. Yet, reliable parameter recovery is essential for interpreting the parameters of a cognitive model. Accordingly, the poor parameter recovery of CT-DDMs prevents the use of parameter estimates to test hypotheses about latent cognitive constructs related to threshold dynamics in decision making (Kruschke and Liddell, 2018). Thus, developing a reliable method for estimating CT-DDM parameters is a crucial step toward understanding how individual differences influence within-trial threshold dynamics. This issue is particularly critical given that the conditions under which individuals adjust their decision thresholds during a single trial are still not well understood.

Since CT-DDMs lack a closed-form likelihood function, most methodological efforts to improve parameter estimation have focused on developing numerical procedures to approximate the likelihood function. One such approach is the integral equation method (e.g., Buonocore et al., 1987, 1990; Smith, 2000; Zhang et al., 2014; Smith and Ratcliff, 2022; Hadian Rasanan et al., 2025), which is both simple and computationally efficient. In this method, the first-passage time distribution of the diffusion process is approximated by solving a linear Volterra integral equation of the second kind. Another prominent approach involves partial differential equations, where the first-passage time distribution is estimated by numerically solving a forward or backward Kolmogorov equation (e.g., Hadian Rasanan et al., 2023, 2024b; Murrow and

Holmes, 2024b [↗](#); Shinn et al., 2020 [↗](#); Boehm et al., 2021 [↗](#); Richter et al., 2023 [↗](#); Voss and Voss, 2008 [↗](#)). In addition to these two numerical methods, which directly estimate the first-passage time distribution, there are simulation-based techniques that approximate the likelihood function using large-scale simulations. These include kernel density estimation approaches (e.g., Turner and Sederberg, 2014 [↗](#); Holmes, 2015 [↗](#)) and neural network-based methods (e.g., Fengler et al., 2021 [↗](#); Radev et al., 2023a [↗](#), 2020 [↗](#)). Each of these likelihood approximation procedures has its strengths and limitations. However, it is important to note that none of these methods fully resolved the reliability issues associated with parameter estimation in CT-DDMs. One reason that more precise likelihood approximation methods did not improve the reliability in parameter estimation is the trade-off between non-decision time and threshold parameters. In other words, there is an identifiability issue in CT-DDMs, which makes estimating threshold dynamics challenging and cannot be resolved by a more precise likelihood approximation. This issue will be discussed further in the next section.

A less explored yet promising solution for improving the reliability of parameter estimation in CT-DDMs is neural-informed cognitive modeling (also known as joint models). Traditionally, these models have been employed to simultaneously fit and predict both behavioral and neural data (e.g., Forstmann and Wagenmakers, 2015 [↗](#); Schall, 2004 [↗](#); Turner et al., 2017 [↗](#), 2015 [↗](#); Nunez et al., 2017 [↗](#); Ghaderi-Kangavari et al., 2022 [↗](#), 2023 [↗](#)). Recent studies have demonstrated that neural-informed approaches can also enhance parameter recovery in EAMs and even render previously unidentifiable parameters identifiable (Nunez et al., 2025 [↗](#); Ghaderi-Kangavari et al., 2023 [↗](#)). Fundamentally, linking DDM parameters to external sources of information — such as neurophysiological signals — introduces additional constraints on the parameter space, thereby enhancing identifiability (e.g., Nunez et al., 2025 [↗](#); Ghaderi-Kangavari et al., 2023 [↗](#)). For instance, Nunez et al. 2025 [↗](#) recently proposed a joint modeling framework to identify the diffusion coefficient (i.e., the noise parameter) in the FT-DDM. By constraining the threshold parameter using external observations (e.g., neural signals), the authors demonstrated that one could simultaneously estimate the drift rate, decision threshold, and diffusion coefficient parameters that cannot be jointly identified using only choice and RT data (Nunez et al., 2025 [↗](#)).

In this work, we explain how non-decision time and collapsing threshold parameters trade off and introduce a non-decision time-informed (NDT-informed) diffusion modeling framework that links non-decision time to external observations, enabling more reliable estimation of collapsing threshold parameters. In contrast to previous methodological studies that primarily focused on likelihood approximation techniques, our contribution lies in proposing a novel *joint modeling* framework, rather than a new estimation algorithm. We demonstrate that informing the model with trial-level noisy measurements of non-decision time — potentially derived from neural data (e.g., Nunez et al., 2019 [↗](#)) — substantially improves parameter recovery in CT-DDMs. In addition, linking non-decision time to external observations enhances the model's fit to behavioral data (i.e., RTs and choices).

The remainder of this paper is organized as follows. First, we present the NDT-informed diffusion modeling framework and discuss a method for estimating non-decision time at the trial level using neural signals, along with theoretical justifications for how constraining non-decision time can improve parameter estimation in CT-DDMs. Next, we report results from an extensive simulation study to assess the effectiveness of the proposed joint modeling framework. Finally, we reanalyze two empirical datasets to illustrate the practical applicability of the method and to examine whether empirical evidence supports the collapsing threshold hypothesis. Finally, we discuss how the proposed NDT-informed diffusion modeling can address the mixed findings in support of CT-DDM and explain how this method can be generalized to other types of diffusion models. We also discuss some behavioral techniques for estimating non-decision time, which do not require neural data.

Non-decision time-informed diffusion modeling

The evidence accumulation process considered in the diffusion models, including CT-DDM, can be represented by a Wiener process (also known as Brownian motion) fluctuating between two time-dependent thresholds (Ratcliff et al., 2016):

$$dX(t) = vdt + sdW(t), \quad b_l(0) < X(0) = x_0 < b_u(0), \quad (1)$$

in which $X(t)$ represents the accumulated evidence until time t , v is the mean evidence accumulation rate (drift rate), s is the diffusion coefficient, x_0 is the starting point bias (also known as pre-decision bias), and $dW(t)$ is the Wiener process. The process starts the accumulation from x_0 and continues until the accumulator crosses either the upper ($b_u(t)$) or the lower ($b_l(t)$) threshold (i.e., $X(t) \geq b_u(t)$ or $X(t) \leq b_l(t)$). v modulates the rate of evidence accumulation, and it corresponds to the quality of the input signal. When the signal-to-noise ratio is high, the task becomes easier, and the speed of evidence accumulation increases. Besides, s determines the level of noise in the process. In addition to these components related to decision time, the DDM also includes a non-decision time parameter (τ), which corresponds to the total time unrelated to the decision process, such as perceptual encoding and motor execution. These — namely, v , x_0 , $b_u(t)$, $b_l(t)$ and τ — are the main components of the model. Figure 1 illustrates a fixed threshold and a hyperbolic collapsing threshold.

Why constraining non-decision time improves parameter estimation in CT-DDMs

A crucial question that needs to be addressed before presenting the NDT-informed diffusion model is why constraining the non-decision time can enhance the parameter estimation of CT-DDMs. To answer this question, we should note that to compute the likelihood function of the CT-DDMs, we need to incorporate the threshold value in the final stopping time (see the discussion on Girsanov change-of-measure theorem in Smith, 2016 and Hadian Rasanan et al. 2025). The final stopping time determines the decision time or, equivalently, is equal to the difference between the RT and the non-decision time (i.e., $t = RT - \tau$). When the decision threshold is fixed over time, its value at the stopping point is the same for any decision time ($RT - \tau$), meaning that non-decision time does not influence the threshold value at the stopping point. However, when the decision threshold is (nonlinearly) time-dependent, the threshold value at the final stopping point depends on non-decision time (i.e., we need to incorporate $b_u(RT - \tau)$ and $b_l(RT - \tau)$). In other words, optimizing the likelihood yields maximizing the probability of reaching a particular threshold value at the final stopping point ($RT - \tau$). Therefore, the estimated non-decision time parameter affects the threshold dynamic by forcing the threshold to reach the value at the stopping time that maximizes the likelihood function. Figure 1 illustrates the effect of non-decision time on the threshold value at the final stopping point.

Consequently, the dependence of the threshold value at the final stopping point on non-decision time causes a trade-off between the threshold parameters (e.g., starting threshold and decay rate) and non-decision time. For instance, if the threshold has a hyperbolic dynamic ($b_u(t) = \frac{\theta}{1+\lambda t}$, in which θ is the starting threshold, and λ is the decay rate), the threshold value at the final stopping point becomes equal to $\frac{\theta}{1+\lambda(RT-\tau)}$. By dividing both the numerator and the denominator by θ , we obtain $\frac{1}{\frac{1-\lambda\tau}{\theta} + \frac{\lambda}{\theta}RT}$, which shows that there is a trade-off between non-decision time, starting threshold, and decay rate. Accordingly, estimating these three parameters simultaneously can be pragmatic. Besides, it implies that inaccurate non-decision time estimation, even if the error is minimal, can lead to erroneous estimates of time-dependent threshold parameters (see Appendix 1 for a simulation study on the tradeoff between non-decision time and starting threshold and

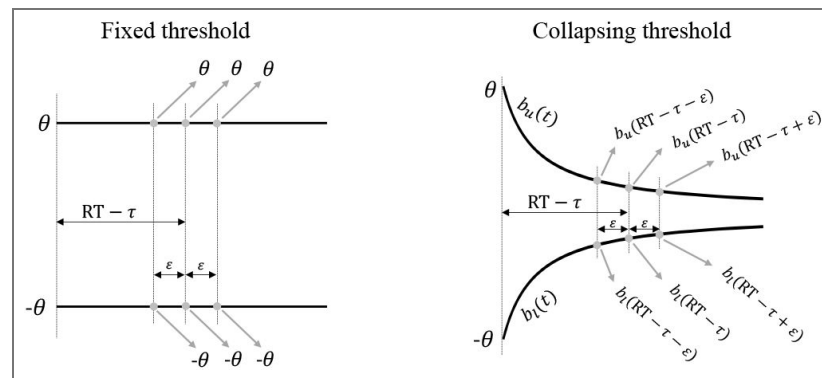


Figure 1. An illustration of the effect of non-decision time on the threshold value on the final stopping point.

The illustration of the evidence accumulation process in the fixed threshold diffusion model (left panel; $b_u(t) = -b_l(t) = \theta$) and hyperbolic collapsing threshold diffusion model (right panel; $b_u(t) = -b_l(t) = \frac{\theta}{1+\lambda t}$).

decay rate). However, constraining non-decision time in the NDT-informed diffusion modeling framework makes all parameters identifiable, thereby improving the reliability of parameter recovery (see also Kira et al., 2025).

Model specification

The traditional diffusion modeling approach estimates model parameters solely based on RT and choice data (e.g., Ratcliff and Rouder, 1998; Ratcliff and Tuerlinckx, 2002; Ratcliff and Smith, 2004). In contrast, in this work, we also assume that an additional source of data related to non-decision time is available. This additional data represents a trial-level noisy measurement of non-decision time, which differs from the actual non-decision time parameter (τ) in the model. Therefore, for each trial n , we assume that a measurement of RT_n , $Choice_n$, and non-decision time measurement (Z_n) are available. It is worth highlighting the difference between Z_n and τ . τ is the mean non-decision time parameter, which is fixed across trials and is estimated through model fitting, whereas Z_n is a trial-level non-decision time measurement (observation) that can be extracted from an external source of data (e.g., from neural signals; see the following section). It is crucial to note that during parameter estimation, we treat Z_n as an input to the model.

Empirical studies on non-decision time measurement usually have reported a right-skewed distribution for their measurements (e.g., Weindel, 2021; Weindel et al., 2025). For instance, the measured perceptual encoding time and motor execution time reported by Weindel et al. 2025 are right-skewed. Therefore, we assume that the observed non-decision time measurements Z_n follow an approximate log-normal distribution, which is right-skewed (we will discuss how this distributional assumption can affect the results later) (It is worth highlighting that this distributional assumption is about the non-decision time measurements and not about across-trial variability in non-decision time). Thus, we model the non-decision time measurements Z_n using a log-normal distribution with parameters μ and σ_z . The available measurements (i.e., observed data) at trial n can be represented as follows:

$$\begin{aligned} RT_n, Choice_n &\sim \text{CT-DDM}(v, x_0, b_u(t), b_l(t), \tau), \\ \log(Z_n) &\sim \mathcal{N}(\mu, \sigma_z^2) \end{aligned} \quad (2)$$

To link the non-decision time measurements (Z_n) to the non-decision time parameter (τ), we assume that the actual non-decision time (τ) is the mean value of the observation distribution, which implies $\tau = \exp(\mu + \sigma_z^2/2)$ or equivalently $\mu = \log(\tau) - \sigma_z^2/2$. Therefore, the joint negative log-likelihood of the NDT-informed CT-DDM can be formulated as follows:

$$\begin{aligned} -LL_{\text{Joint}}(RT, Choice, Z) = & -LL_{\text{CT-DDM}}(RT, Choice \\ & | v, x_0, b_u(t), b_l(t), \tau, \sigma_z) = \\ & -\sum_n^N \left[\frac{(\log(Z_n) - \log(\tau) - \sigma_z^2/2)^2}{2\sigma_z^2} \right. \\ & \left. + \frac{1}{2} \log(2\pi\sigma_z^2 Z_n^2) \right]. \end{aligned} \quad (3)$$

The first term ($LL_{\text{CT-DDM}}$) corresponds to the likelihood of the choice behavior (i.e., RT and choice data) predicted by the CT-DDM. The second term represents the log-likelihood of the observed non-decision times, which are assumed to follow a log-normal distribution with a mean of τ and a shape parameter of σ_z . The non-decision time parameter τ is shared across both terms, and linking it to trial-level measurements of non-decision time (Z_n) helps constrain its estimation. By minimizing this joint negative log-likelihood function, we can simultaneously estimate the CT-DDM parameters that predict human choice behavior and the summary statistics of the non-decision time observations.

As previously noted, an exact likelihood function for CT-DDMs is not available. Nevertheless, several numerical methods exist for approximating this likelihood, including the integral equation method (e.g., [Smith, 2000](#); [Zhang et al., 2014](#); [Smith and Ratcliff, 2022](#)) and the partial differential equation method (e.g., [Hadian Rasanan et al., 2023](#); [Shinn et al., 2020](#); [Boehm et al., 2021](#); [Richter et al., 2023](#)). Any of these methods can be used to compute $-LL_{CT-DDM}$. In this work, we employed the integral equation method, which is a relatively simple and accurate approach ([Richter et al., 2023](#); [Hadian Rasanan et al., 2025](#)). In Appendix 2, we detail the procedure for computing $-LL_{CT-DDM}$ using the integral equation method. Moreover, we replicated the main results using a BayesFlow approach ([Radev et al., 2020](#)) to assess sensitivity to the likelihood approximation method.

How to measure non-decision time

The central assumption of the NDT-informed diffusion modeling is that the non-decision time measurement (Z_n) is available for each trial. Thus, providing a reliable measurement of non-decision time is crucial for the proposed model. Measuring non-decision time using neurophysiological signals like electromyographic (EMG) or electroencephalogram (EEG) has been studied in the field of model-based cognitive neuroscience for a decade (e.g., [Servant et al., 2016](#); [Nunez et al., 2019](#); [Weindel et al., 2021a](#); [Ghaderi-Kangavari et al., 2022](#)). Accordingly, several techniques have been proposed in the literature to measure non-decision time. One class of methods involves measuring single-trial event-related potentials (ERPs), enabling researchers to link the ERP components to cognitive parameters of the diffusion model through a (non-)linear regressor (e.g., [Nunez et al., 2017](#); [Bridwell et al., 2018](#); [Ghaderi-Kangavari et al., 2022](#), [2023](#)). One main single-trial ERP component of interest is the N200 peak latency, which is thought to correlate with perceptual encoding time ([Nunez et al., 2019](#)). However, single-trial N200 latencies capture only a portion of non-decision time, which relates to perceptual encoding, and thus, cannot be used directly as a measurement of total non-decision time. Furthermore, incorporating the N200 component as a regressor for the non-decision time parameter does not resolve the identifiability issue of the CT-DDM. This limitation was highlighted by [Ghaderi-Kangavari et al. 2023](#), who conducted a systematic study on parameter recovery for various neural-informed diffusion models. They reported unreliable parameter estimates for the CT-DDM in which non-decision time was linked to trial-level N200 latency via a regressor (see models 5 and 6 in [Ghaderi-Kangavari et al., 2023](#)). The poor parameter recovery in their models arises from a trade-off between the threshold parameters and the regression coefficients — a problem similar to the identifiability issue we discussed earlier. Therefore, informing the model with N200 latency does not resolve the underlying identifiability problem.

Similarly, an EMG signal can be used to measure the execution time of motor responses ([Weindel et al., 2021a](#)). In this approach, we can detect and measure the motor execution time based on the muscle activity. Although this method provides an accurate measure of motor execution, it does not capture total non-decision time and fails to address the parameter recovery issues in CT-DDMs reported by [Ghaderi-Kangavari et al. 2023](#). Furthermore, recent evidence shows that at least some portion of the EMG signal is related to evidence accumulation ([Servant et al., 2021](#); [Weindel, 2021](#); [Weindel et al., 2021b](#)). Although a potential approach would be to use both EMG and EEG signals together to estimate perceptual encoding and motor execution time, this approach is relatively complicated (see also [Kelly et al. 2021](#) for a discussion on how to extract perceptual encoding onset, evidence accumulation onset, and post-decision motor-execution time from EEG signals).

More recent advancements in neural signal analysis have led to the development of methods to discover single-trial EEG or MEG events across the entire trial-level time series (e.g., [Anderson et al., 2016](#); [Weindel et al., 2024](#)). These methods provide the sequence and timing of latent cognitive steps at each trial and have already been used to uncover the cognitive stages in various empirical paradigms, including perceptual decision tasks ([Van Maanen et al., 2021](#)), lexical decision tasks ([Berbery et al., 2021](#); [Krause et al., 2024](#)), and associative recognition tasks ([Van Maanen et al., 2021](#)). In this paper, we have employed the hidden multivariate pattern

(HMP) method (Weindel et al., 2024) to extract non-decision time from the EEG signal. It has been demonstrated that this method can decompose the RT into components associated with decision-relevant features versus decision-unrelated physical properties of the stimuli (Weindel et al., 2025). Thus, we can extract the decision time from neural signals, and by subtracting the estimated decision time from observed RT, we can obtain a measurement of non-decision time in each trial. The entire procedure of NDT-informed diffusion modeling, which involves extracting non-decision time from neural signals using the HMP method, is presented in Figure 2.

Simulation Study 1: Parameter recovery

Methods

We conducted a parameter recovery study using simulated data to evaluate whether the parameters of the NDT-informed CT-DDM can be estimated reliably. To assess the robustness of our method, we considered two distinct nonlinear functional forms for the threshold dynamics. Theoretical studies on optimality have shown that the optimal decision threshold follows a nonlinear, monotonically decreasing trajectory (Fudenberg et al., 2018; Frazier and Yu, 2007). These studies suggest that hyperbolic or exponential functions are the most plausible candidates for modeling the optimal threshold (Fudenberg et al., 2018; Frazier and Yu, 2007). These collapsing thresholds have already been used in several studies (e.g., Olschewski et al., 2025; Bhui, 2019; Milosavlje-vic et al., 2010; Voskuilen et al., 2016; Hanks et al., 2011). Additionally, Smith and Ratcliff, 2022 demonstrated that an urgency gating model with a linear urgency signal is mathematically equivalent to a hyperbolic CT-DDM (see also Trueblood et al., 2021). It is also important to note that prior research has reported poor parameter recovery for nonlinear collapsing threshold models (Evans et al., 2020b; Murrow and Holmes, 2024a). Therefore, we selected the exponential collapsing threshold diffusion model (ECT-DDM) and the hyperbolic collapsing threshold diffusion model (HCT-DDM) — two of the most theoretically motivated and widely supported nonlinear threshold dynamics — for our simulations (some studies have also employed the Weibull function for modeling threshold dynamics [e.g., Hawkins et al., 2015; Ging-Jehli et al., 2025; Evans et al., 2020a]). This choice is typically made for its flexibility, allowing the model to mimic a variety of threshold dynamical forms. However, this function has three parameters, which causes some trade-off among the threshold parameters and makes the parameter recovery much more difficult):

$$\text{Hyperbolic : } b_u(t) = -b_l(t) = \frac{\theta}{1+\lambda t} \quad (4)$$

$$\text{Exponential : } b_u(t) = -b_l(t) = \theta \times \exp(-\lambda t) \quad (5)$$

where θ is the starting threshold (i.e., $b_u(0) = \theta$) and $\lambda > 0$ is the decay rate. $b_u(t)$ stands for the upper threshold, and the lower threshold is equal to the reflection of the upper threshold (i.e., $b_l(t) = -b_u(t)$). In addition to these two non-linear collapsing threshold dynamics, we also considered linear collapsing threshold models, defined as $b_u(t) = -b_l(t) = \theta - \lambda t$, to check the robustness of the results. The results were largely consistent with the non-linear models, as detailed in Appendix 3. In all simulations, we assumed that the accumulation process starts from zero (i.e., starting point $x_0 = 0$), which implies that there is no *a priori* bias towards one option in the simulation data. Also, we set the diffusion coefficient equal to one (i.e., $s = 1$). To simulate RT and choice, we considered the following discrete form of the accumulation process (1) with a time step $\Delta t = 0.001$:

$$X(t + \Delta t) = X(t) + v\Delta t + \sqrt{\Delta t}e, \quad e \sim \mathcal{N}(0, 1), \text{ and } X(0) = 0. \quad (6)$$

To cover all possible model behavior in the simulations, we used the following uniform distributions for the CT-DDM parameters (i.e., δ , τ , θ , and λ) and the shape parameter of the non-decision time measurements (σ_z) ($\mathcal{U}[a, b]$ represent a uniform distribution over the interval between a and b):

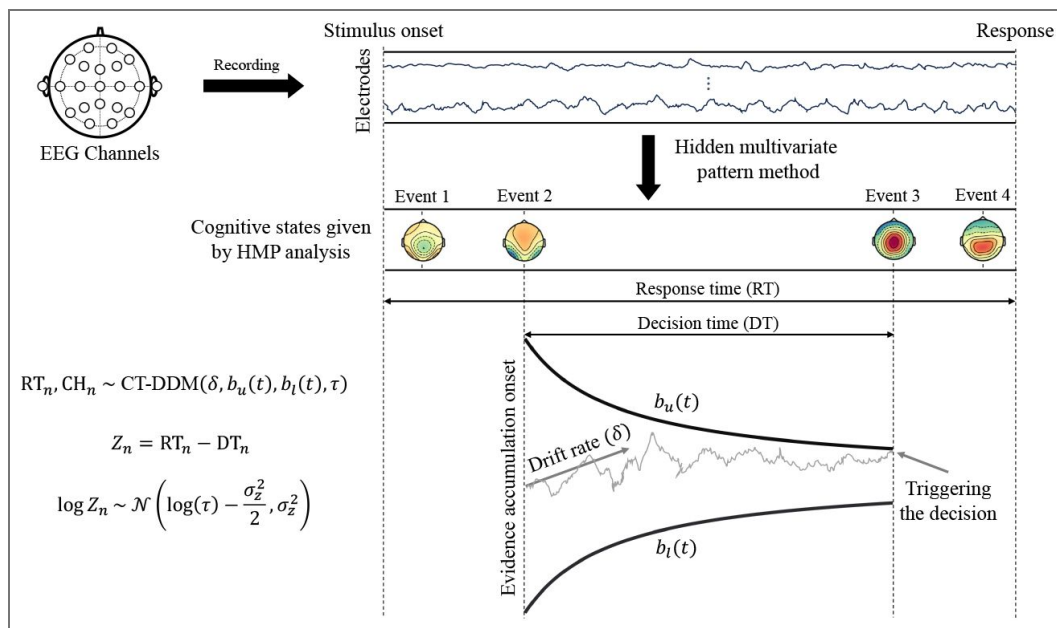


Figure 2. Illustration of the non-decision time-informed diffusion model.

The HMP analysis method extracts the timing of cognitive states in each trial from EEG signals. Then, the decision time is considered the duration between the cognitive states that determines the decision process, such as the N200 latency, which marks the end of perceptual encoding (Nunez et al., 2019), and the peak of central parietal positivity, which indicates the end of the decision process (O’Connell et al., 2012 — see also Weindel et al. 2025 for a detailed discussion). Subtracting the decision time from the response time gives an approximation of non-decision time in each trial. The extracted non-decision time measurements are employed to constrain the non-decision time parameter in the CT-DDM.

$$\begin{aligned}\theta &\sim \mathcal{U}[1.5, 4], \\ \lambda &\sim \mathcal{U}[0.1, 2], \\ \delta &\sim \mathcal{U}[0, 3], \\ \tau &\sim \mathcal{U}[0.05, 1], \\ \sigma_z &\sim \mathcal{U}[0.1, 1].\end{aligned}$$

After sampling random parameters from the uniform distributions described above, we generated RT, choice, and random non-decision time measurements (Z_n) for various trial numbers (i.e., 100, 250, 500, 750, and 1000). For each number of trials, we generated 1000 datasets. As we already mentioned, we assumed that the actual non-decision time (τ) is assumed to be equal to the mean value of the log-normally distributed non-decision time measurements. Thus, the location parameter of the log-normal distribution is obtained by $\mu = \log(\tau) - \frac{\sigma_z^2}{2}$. Therefore, in each trial, we sampled a noisy observation of non-decision time using the following log-normal distribution:

$$\log(Z_n) \sim \mathcal{N}\left(\log(\tau) - \frac{\sigma_z^2}{2}, \sigma_z^2\right)$$

We have employed the differential evolution optimization routine to minimize the joint negative log-likelihood (3). Also, to approximate the likelihood function of the CT-DDM, we used the integral equation method with a time step $\Delta t = 0.02$ (see Appendix 2 for the details of likelihood approximation). A problem with the maximum likelihood procedure is its sensitivity to outliers, which can yield likelihood values of zero (Ratcliff and Tuerlinckx, 2002). Outliers can affect the fitting procedure and cause convergence problems (Ratcliff and Tuerlinckx, 2002; Heathcote et al., 2002). We truncated the likelihood values lower than 10^{-14} to address this issue. To account for potential influences of the estimation method on parameter inference, we replicated the main results using amortized Bayesian inference as implemented in BayesFlow (Radev et al., 2020, 2023b). The obtained results based on this alternative approach are presented in Appendix 4.

To evaluate the precision of parameter recovery, we employed the R-squared (R^2) index, which is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (\vartheta_i - \hat{\vartheta}_i)^2}{\sum_{i=1}^N (\vartheta_i - \bar{\vartheta})^2}$$

in which ϑ and $\hat{\vartheta}$ stand for the actual generating parameter and estimated parameter, respectively, and $\bar{\vartheta}$ in the denominator shows the mean of the true parameter. Based on the R^2 index, we can assess how well the estimated parameters align with the actual data-generating parameters. This metric has already been employed in the literature to evaluate parameter estimation methods and assessment of estimation reliability in cognitive models (e.g., Radev et al., 2020; Hadian Rasanan et al., 2025).

Results

Figure 3 illustrates the R^2 value for the threshold parameters (i.e., starting threshold θ , and decay rate λ) for both the NDT-informed and uninformed CT-DDM, based on 500 trials. First, this plot demonstrates a substantial improvement in the reliability of threshold parameter recovery in the NDT-informed CT-DDM compared to the uninformed CT-DDM. This improvement is more pronounced for the hyperbolic threshold than the exponential threshold. Second, the results indicate that threshold parameter estimation is generally reliable for both the hyperbolic and exponential models, with particularly high precision observed in the exponential case.

Figure 4 shows the estimated parameters against the true parameters for both NDT-informed CT-DDMs and uninformed CT-DDMs based on 500 trials. The results suggest that the starting threshold parameter in both hyperbolic and exponential uninformed CT-DDM is unrecoverable, which aligns with previous parameter recovery assessment studies (Evans et al., 2020b; Murrow and Holmes, 2024a; Ghaderi-Kangavari et al., 2023). Additionally, the decay rate in the hyperbolic uninformed CT-DDM exhibits poor parameter recovery. In contrast, the parameters estimated by the NDT-informed CT-DDM closely match the true data-generating parameters. The

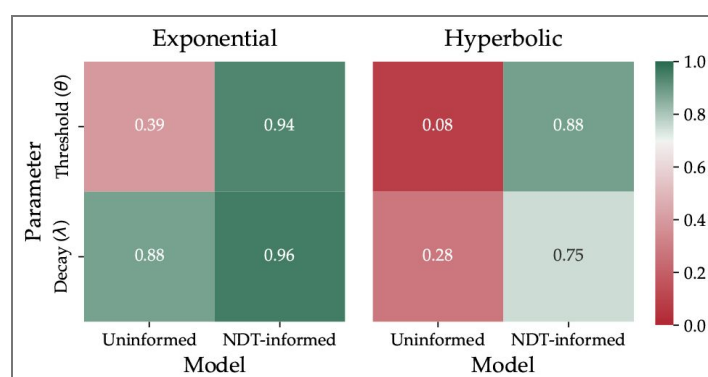


Figure 3. R^2 values measuring the agreement between estimated and ground truth threshold parameters in exponential (left) and hyperbolic (right) collapsing threshold models.

only challenging parameter to estimate is the decay rate in the hyperbolic CT-DDM. As shown in Figures 3 and 4, recovering this parameter is more challenging in the hyperbolic model compared to the exponential CT-DDM.

Figure 5 illustrates the sensitivity of parameter estimation in both NDT-informed and uninformed CT-DDMs as a function of the number of trials. As expected, increasing the number of trials generally improves parameter recovery. However, in the uninformed approach, the collapsing threshold parameters cannot be reliably estimated even with 1000 trials. In contrast, NDT-informed CT-DDMs achieve reliable parameter recovery with substantially fewer trials. Specifically, in the NDT-informed hyperbolic model, accurate estimation of the decay rate requires more than 250 trials. In the exponential NDT-informed model, both the starting threshold and decay rate can be reliably estimated with as few as 100 trials. The drift rate and non-decision time parameters are estimated with high precision, even with 100 trials in both informed and uninformed models, regardless of whether the threshold follows a hyperbolic or exponential form.

Figure 6 presents the sensitivity of the parameter estimation in NDT-informed CT-DDM to the noise level in non-decision time observation (Z_n). We simulated data using three fixed levels of standard deviation for the non-decision time measurements (i.e., SD = 0.3, 0.6, and 0.9 seconds), with 500 trials generated for each noise level. The results demonstrate that even with a high level of variability (noise) in non-decision time measurements, parameter estimation remains reliable. Moreover, these results demonstrate that the parameter recovery of NDT-informed CT-DDMs is better than the parameter estimation of the uninformed CT-DDM, even under high levels of noise in non-decision time measurements.

Additionally, we examined the effect of systematic bias in non-decision time measurement on parameter recovery. Appendix 5 presents the results in the presence of biased non-decision time measurement (i.e., systematically underestimated or overestimated). The results revealed that, even in the presence of biased non-decision time measurement, the actual generating parameters show high correlations with the estimated parameters. Underestimation in non-decision time leads to overestimation in the starting threshold and decay rate. Conversely, overestimation in non-decision time leads to underestimation in both the starting threshold and non-decision time.

Simulation Study 2: Model recovery

Methods

One of the main counterarguments against CT-DDMs is that the slow error responses can also be explained by across-trial variability in drift rate (Voskuilen et al., 2016). An FT-DDM that incorporates across-trial variability in drift rate can account for slow errors, and it often achieves fitting performance comparable to that of the CT-DDM when applied to empirical data (Voskuilen et al., 2016; Smith and Ratcliff, 2022). This similarity in model performance makes it difficult to determine, based solely on model fit, whether the underlying evidence accumulation process involves a collapsing threshold or a fixed threshold with drift variability. The aim of this section is to show that NDT-informed diffusion modeling enables distinguishing between the CT-DDMs and FT-DDMs with variability in drift rate. To this aim, we conducted a cross-model fitting simulation and a model recovery simulation.

The cross-model fitting simulation is conducted to investigate how parameter estimation in CT-DDMs is affected when the true generative process is an FT-DDM with across-trial variability in the drift rate. In this simulation, we generated data from an FT-DDM with across-trial variability in drift rate and then fitted a CT-DDM without trial-to-trial variability to the simulated data. Across-trial variability in drift rate was modeled as a normal distribution with mean δ and standard deviation of η (i.e., $v_t \sim \mathcal{N}(\delta, \eta)$, in which v_t is the trial-level drift rate). The following uniform distributions were used to sample the parameters of the FT-DDM:

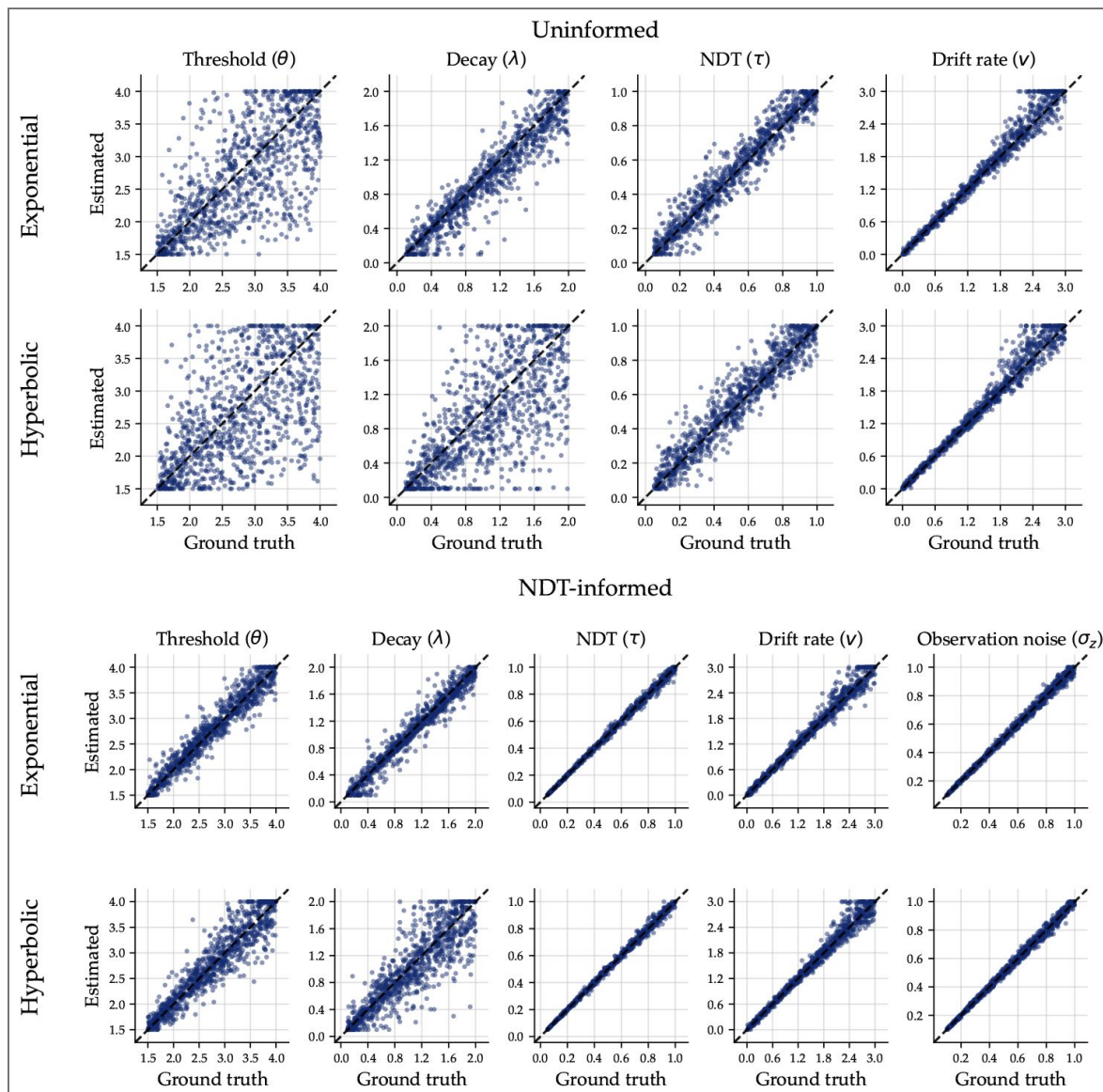


Figure 4. Estimated versus ground truth parameter values for two modeling approaches.

The top two rows show parameter recovery for models estimated without additional constraints (“Uninformed”), while the bottom two rows show recovery for joint models that incorporate additional non-decision time observations (“NDT-informed”). Each point represents a recovered parameter estimate plotted against its corresponding true data-generating value. Rows correspond to different functional forms of threshold (i.e., Exponential and Hyperbolic). Dashed diagonal lines indicate perfect recovery.

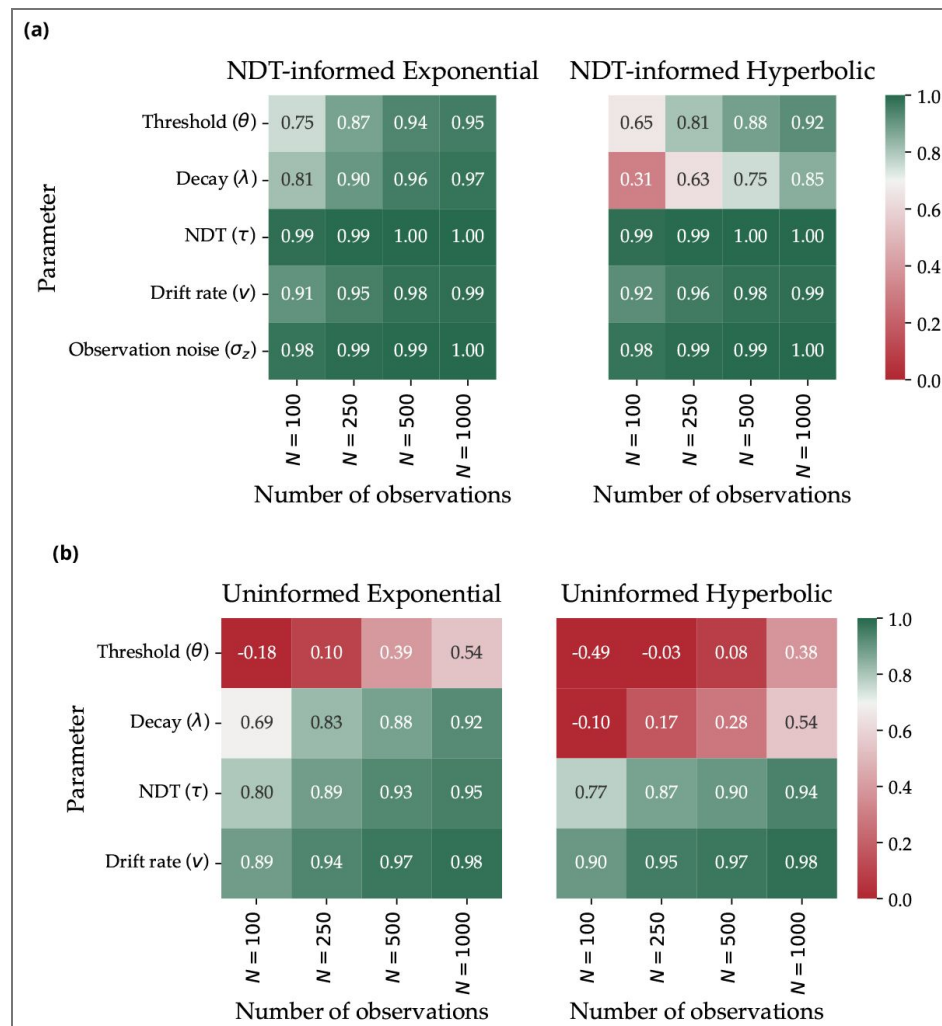


Figure 5. Illustration of the sensitivity of parameter estimation to the number of trials in the NDT-informed (a) and uninformed models (b).

R^2 values measure the agreement between estimated and ground truth parameters of the exponential (left) and hyperbolic (right) collapsing threshold models.

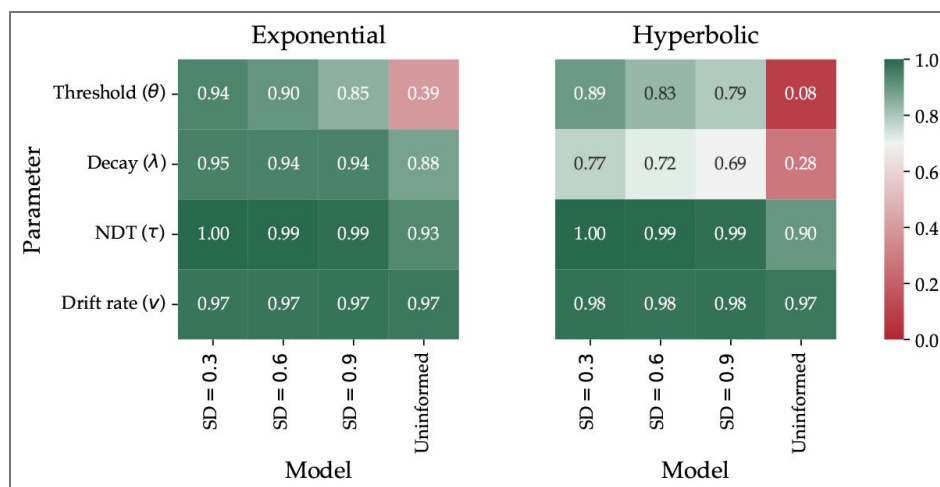


Figure 6. Illustration of the sensitivity of parameter estimation to the noise level in the non-decision time observations.

R^2 values measuring the agreement between estimated and ground truth parameters of the exponential (left) and hyperbolic (right) collapsing threshold models.

$$\begin{aligned}\theta &\sim \mathcal{U}[1.5, 4], \\ \eta &\sim \mathcal{U}[0.1, 1], \\ v &\sim \mathcal{U}[0, 3], \\ \tau &\sim \mathcal{U}[0.05, 1], \\ \sigma_z &\sim \mathcal{U}[0.1, 1]\end{aligned}$$

We generated 1000 parameter sets, and for each set, we simulated 500 trials from an FT-DDM with across-trial variability in drift rate. As in the previous simulation study, we assumed that trial-level measurements of non-decision time are available, enabling us to fit NDT-informed diffusion models. Specifically, we fit both an NDT-informed CT-DDM with a hyperbolic collapsing threshold and an NDT-informed CT-DDM with an exponential collapsing threshold to each simulated dataset.

In the model recovery study, we considered two classification problems: a) FT-DDM with drift variability versus exponential CT-DDM and b) FT-DDM with drift variability versus hyperbolic CT-DDM. Specifically, in each classification study, we generated 1000 parameter sets — 500 from an FT-DDM with across-trial variability in drift rate and 500 from the CT-DDM — and simulated 500 trials from each generative model. In each classification study, both the FT-DDM with drift variability and the CT-DDM were then fitted to every simulated dataset. To infer the underlying generating process, we considered two approaches. In one approach, we only considered how well each model was fitted to the simulated data sets, and we compared the goodness-of-fit measures to infer the underlying generating process (as the number of parameters in both considered models is identical [i.e., the complexity of the two models is the same], we compared the log-likelihood value as a measure of goodness-of-fit to classify the underlying generative model). In another approach, we also considered the recovered decay rate (λ) value and used this parameter as an additional indicator for the underlying generative process. Therefore, in the second approach, on top of comparing the goodness-of-fits, if the recovered decay rate was lower than 0.1, we classified it as an FT-DDM model.

Results

Figure 7 [illustrates](#) the parameter recovery results for the model misspecification study (please note that, as the actual generating decay rate value for the FT-DDM is always zero, we employed a density plot to illustrate the density of the estimated decay rate). The most important result is related to the decay rate parameter (λ). For more than 75% of the datasets, the estimated decay rate is approximately zero, and only for 5% is the estimated decay rate larger than 0.1. Moreover, these results demonstrate that both the non-decision time parameter and the variability of non-decision time observations can be recovered with high accuracy. The recovered drift rate and starting threshold parameters showed a slight underestimation for larger values. However, the R^2 value remains high for these parameters (i.e., $R^2 > 0.85$).

Table 1 [summarizes](#) the model recovery results for both exponential and hyperbolic CT-DDMs. The results indicate that the exponential model exhibits better recovery performance than the hyperbolic model, consistent with its superior parameter recovery. More importantly, incorporating the estimated decay rate into model inference substantially improves model recovery accuracy. In other words, rather than relying solely on goodness-of-fit indices, using the estimated decay rate to guide model classification (i.e., models with an estimated decay rate below 0.1 were classified as FT-DDMs) can improve the precision of model recovery. Because the NDT-informed framework provides reliable estimates of collapsing threshold parameters, these parameter values can be used confidently in model inference, thereby improving the precision of model recovery (however, we should also note that model recovery is highly sensitive to the experimental design considered. For instance, the experimental designs that include a condition with a high level of difficulty [i.e., the drift rate close to zero], then these models are more distinguishable).

In sum, the simulation results confirm that incorporating non-decision time information into the diffusion model enables distinguishing between CT-DDM and FT-DDM. Moreover, these results suggest that NDT-informed diffusion modeling can be reliably applied to infer the underlying threshold dynamics. In other words, if the underlying generating process is an FT-DDM (with or

Figure 7. Estimated versus true parameter values for cross-fitting of CT-DDM on FT-DDM with across-trial variability in drift rate.

In each panel except the rightmost one, each point represents a recovered parameter estimate plotted against its corresponding true data-generating value, and the dashed diagonal line indicates perfect recovery. The right column illustrates the density of the estimated decay rate parameter (λ). Rows correspond to different functional forms of threshold (i.e., Exponential and Hyperbolic).

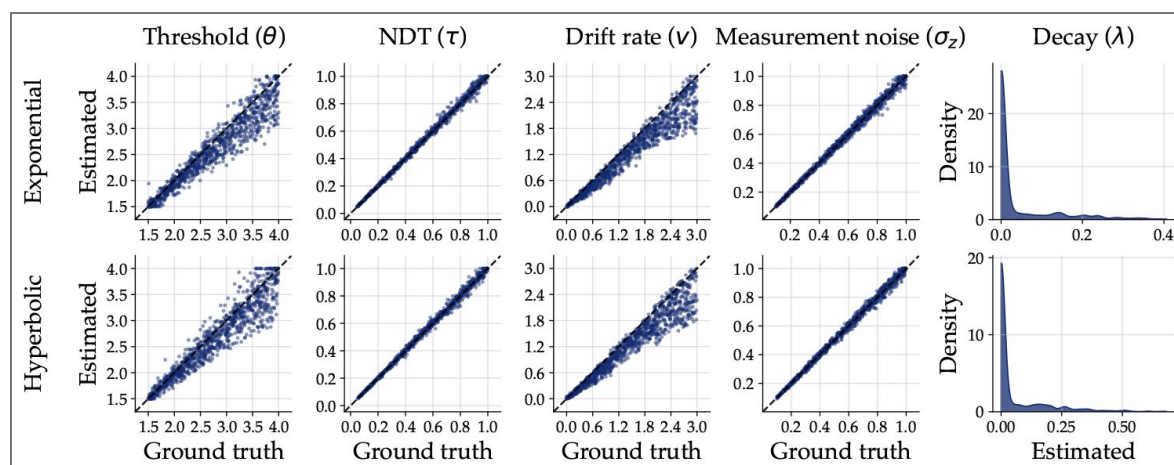


Table 1. The model recovery results for exponential and hyperbolic collapsing threshold diffusion models.

	Exponential	Hyperbolic
Goodness of fit only	89.3%	86.9%
Goodness of fit and decay value	97.7%	95.1%

without across-trial variability), then the estimated decay rate would be very close to zero or a relatively very small value. Therefore, we can infer the underlying threshold dynamics through the estimated decay rate value in the NDT-informed diffusion modeling framework.

Applications to empirical data

We present two case studies to: (1) show the feasibility of the proposed NDT-informed diffusion modeling approach, (2) explore whether informing the diffusion model with noisy measurements of non-decision time can enhance model fitting to behavioral data, and (3) investigate whether people adjust the decision threshold within a single trial (at least in the case studies considered here) (the third aim is set due to the controversial findings in the literature that did not find evidence in support of CT-DDMs [e.g., Voskuilen et al., 2016 [↗](#); Smith and Ratcliff, 2022 [↗](#); Milosavljevic et al., 2010 [↗](#); Karşilar et al., 2014 [↗](#)]). To analyze the behavioral data, we consider six different diffusion models. For each dataset, we fitted the NDT-informed HCT-DDM, ECT-DDM, and FT-DDM to compare the fixed threshold against the collapsing threshold using the joint modeling framework. Also, to check for improvements of NDT-informed, we fitted the uninformed version of HCT-DDM and ECT-DDM. Additionally, as a benchmark, we fitted an FT-DDM that includes trial-to-trial variability in drift rate. Then, we employed the Bayesian information criterion (BIC) to evaluate the goodness-of-fit and to compare the models. We compared the considered diffusion models in two ways: with respect to fitting on behavioral data (only RT and choice) and also fitting on all observations, including RT, choice, and non-decision time measurements. Thus, we defined these two BICs for model comparisons:

$$\begin{aligned} \text{BIC}_{\text{Behavior}} &= -2 \times LL_{\text{CT-DDM}} + K_{\text{CT-DDM}} \times \log(N_{\text{Observation}}), \\ \text{BIC}_{\text{Joint}} &= -2 \times LL_{\text{Joint}} + K_{\text{Joint}} \times \log(N_{\text{Observation}}), \end{aligned}$$

in which $LL_{\text{CT-DDM}}$ is the log-likelihood of the (collapsing threshold) diffusion model (see Appendix 2 for the details of the likelihood approximation using the integral equation method), LL_{Joint} is the joint log-likelihood of behavioral data and non-decision time estimates defined in equation 3 [↗](#), $K_{\text{CT-DDM}}$ is the number of free parameters in the (collapsing threshold) diffusion model, K_{Joint} is the number of free parameters in the joint model, and $N_{\text{Observation}}$ is the number of observed data points. $\text{BIC}_{\text{Behavior}}$ only considers the goodness-of-fit on RT and choice, while $\text{BIC}_{\text{Joint}}$ also considers how well the model can predict the non-decision time observations. Thus, we can compare all six considered models based on $\text{BIC}_{\text{Behavior}}$ and check whether constraining the non-decision time can improve the fitting of behavioral data. However, $\text{BIC}_{\text{Joint}}$ is only used for comparing the NDT-informed diffusion models. It is worth noting that comparing joint and behavioral models using $\text{BIC}_{\text{Behavior}}$ is uncommon and constitutes a limitation of the model comparison study.

Study 1: Weindel et al., 2025 [↗](#) dataset

In the first study, we considered data from Weindel et al. 2025 [↗](#). Twenty-six participants took part in the study and performed a binary choice task involving contrast discrimination. On each trial, two Gabor patches were presented, and participants indicated which option had the higher contrast. The difference in contrast level between the two options was always fixed at 5%, while the absolute contrast levels varied from 5% to 95%. In total, participants completed eight experimental blocks, each containing 140 trials. At the beginning of each block, participants received a speed-accuracy instruction. In half of the blocks, participants were requested to focus on decision speed, and in the other half, they were asked to focus on accuracy. The decision difficulty was designed to follow Fechner's law (Fechner, 1860 [↗](#)). According to this law, keeping the difference between options constant makes it harder to discriminate between two options as the overall value increases. To easily approximate this law, the following relation between contrast level and drift rate was implemented:

$$v = v_0 + v_1 \times \log(\text{Contrast})$$

in which v_0 , and v_1 are free parameters. v_0 is the baseline drift rate value, and v_1 modulates the sensitivity to the contrast level. The non-decision time measurements were obtained from the fit of an HMP model using the standard cumulative procedure (Weindel et al., 2024) on preprocessed EEG data with minimal trial rejection (see Weindel et al., 2025, for the detailed procedure). The HMP model estimated five cognitive events. Their associated single-trial timing and electrode contributions allowed us to interpret these events as relating to early visual processing, perceptual encoding, attention orientation, motor planning, and decision termination. For this analysis, we subtracted the time between the attention orientation (third event) and decision termination (last event) from the RT at each trial to yield a measure of non-decision time. See Figure 8 for the sequence of these events and the time interval between them.

The quantitative results revealed that NDT-informed HCT-DDM outperforms the other models. Table 2 presents the mean estimated parameters of the computational models and a model comparison result based on BIC. The best-fitting model in both speed and accuracy conditions is the NDT-informed HCT-DDM. Also, these results show that both NDT-informed models have better predictive accuracy on RT and choice compared to the two uninformed counterpart models (i.e., based on $\text{BIC}_{\text{Behavior}}$). The estimated parameters in both NDT-informed models show that the mean starting threshold in the speed condition is lower than in the accuracy condition. However, the mean decay rate is larger in the speed condition compared to the accuracy condition. This implies that the decision threshold starts at a lower point and declines faster in the speed condition than in the accuracy condition.

Figure 9 demonstrates the estimated threshold dynamics for each individual (gray lines) and the group average (blue lines) by NDT-informed CT-DDMs (see Appendix 6 for the estimated collapsing threshold by uninformed models). The estimated threshold for each individual indicates that most of the participants employed a collapsing threshold strategy (even in the accuracy condition).

Figure 10 shows the predictions of the NDT-informed models for both the speed and accuracy conditions. The FT-DDM significantly overestimates the later RT quantile for correct and incorrect responses, in both speed and accuracy conditions. This overestimation is more pronounced in the speed condition. Moreover, FT-DDM underestimates the proportion of incorrect responses. In contrast to FT-DDM, which cannot accurately predict empirical data, particularly in the speed condition, the NDT-informed CT-DDMs exhibit high predictive accuracy.

In sum, both quantitative and qualitative results revealed that the NDT-informed models perform better than the uninformed models in this study. In other words, constraining CT-DDM with non-decision time improved both the model fits and predictions. In addition to improving model fitting, the results obtained in this study also support the collapsing threshold against the fixed threshold. First, the mean estimated decay rates reported in Table 2 are significantly large. Second, all the considered CT-DDMs have a lower BIC and also better posterior prediction compared to the FT-DDM, which includes across-trial variability in drift rate (see the Appendix 7 for the results of FT-DDM). Third, the estimated threshold dynamics in Figure 9 indicate that most participants adopted a collapsing threshold. Therefore, considering all the results together, this study highlights the advantage of the NDT-informed modeling approach and supports CT-DDMs.

Study 2: Boehm et al., 2014 dataset

The dataset for the second case study is taken from Boehm et al. 2014. In this experiment, twenty-five participants performed a random-dot motion task with a speed-accuracy trade-off manipulation. Before each trial, a cue (i.e., “AC” to emphasize accuracy in the subsequent trial or “SP” to emphasize speed in the following trial) was presented to inform participants how they should decide in the next trial. On each trial, participants viewed a cloud of moving dots for 1.5 s,

Figure 8. The event sequence and timing of each event estimated by the HMP method averaged over all participants (Weindel et al., 2025 [DOI](#)).

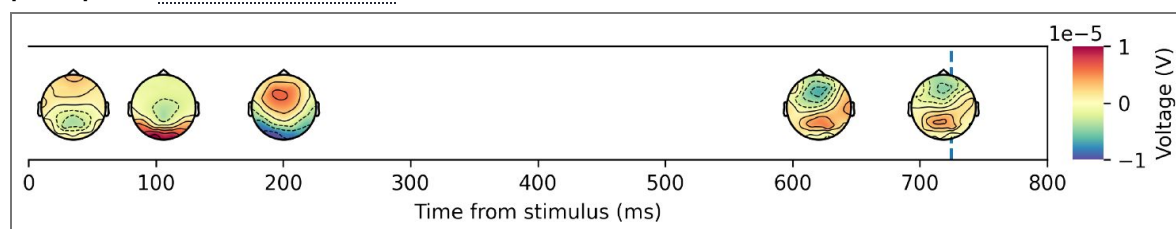


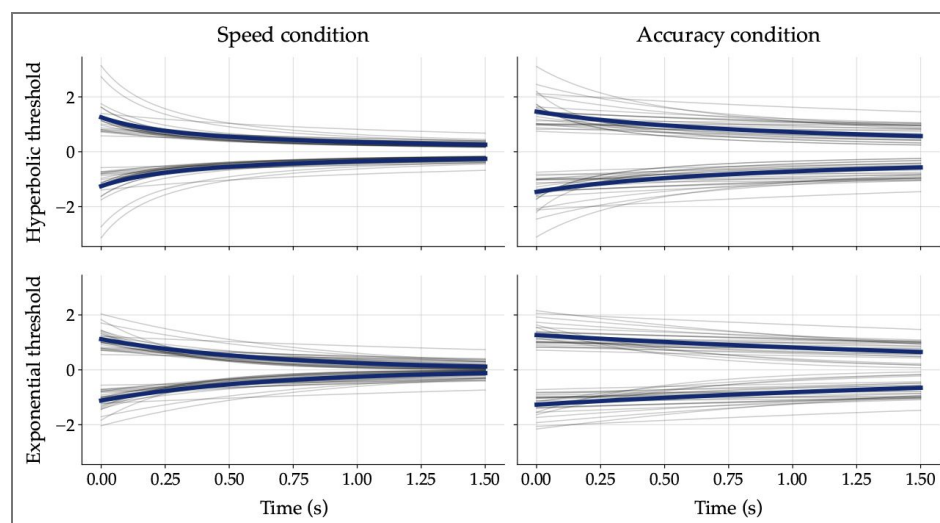
Table 2. The mean estimated parameters and goodness of fit results for Study 1.

	Model	v_0	v_1	τ	θ	λ	σ_z	BIC _{Joint}	BIC _{Behavior}
Speed	NDT-inf HCT-DDM	2.77	0.51	0.19	1.28	2.67	0.31	-36990.16	2644.03
	NDT-inf ECT-DDM	2.80	0.51	0.19	1.16	1.61	0.31	-36952.86	2751.98
	NDT-inf FT-DDM	3.01	0.54	0.21	0.69	—	0.37	-29207.05	6528.17
	Uninf HCT-DDM	2.87	0.53	0.21	1.16	2.73	—	—	2707.24
	Uninf ECT-DDM	2.92	0.54	0.21	1.07	1.40	—	—	2802.94
Accuracy	NDT-inf HCT-DDM	3.17	0.59	0.21	1.64	1.26	0.33	-18225.91	14919.20
	NDT-inf ECT-DDM	3.19	0.59	0.21	1.40	0.58	0.33	-17862.53	15153.47
	NDT-inf FT-DDM	3.11	0.57	0.24	1.00	—	0.41	-9196.26	17301.73
	Uninf HCT-DDM	3.32	0.63	0.28	1.47	1.09	—	—	15863.86
	Uninf ECT-DDM	3.44	0.65	0.31	1.07	0.32	—	—	15915.51

Note. NDT-inf = NDT-informed and Uninf = Uninformed

Figure 9. The estimated threshold dynamics from NDT-informed CT-DDMs for individuals in Study 1.

The first row illustrates the estimated hyperbolic threshold dynamics, and the second row illustrates the estimated exponential threshold dynamics. The left column shows the threshold dynamics in the speed condition, and the right column shows the accuracy condition. Each gray line corresponds to a subject. The blue line corresponds to the average group level.



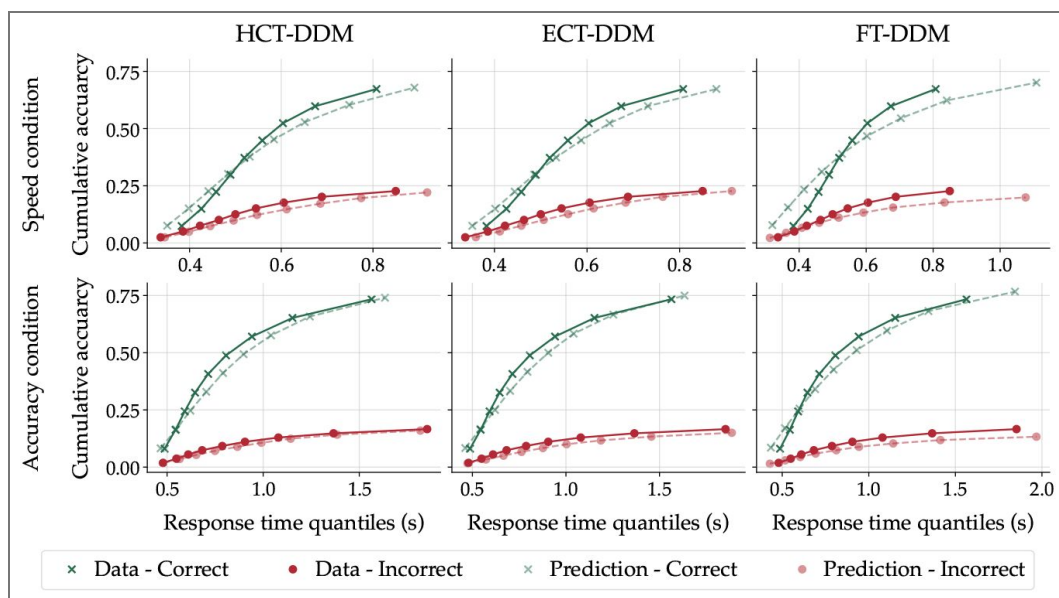


Figure 10. Prediction of NDT-informed models against empirical data for speed (top row) and accuracy (bottom row) conditions in Study 1.

In each panel, the x-axis represents the response time quantiles in seconds, and the y-axis represents the cumulative choice proportion.

of which a subset moved coherently in one direction while the rest moved randomly, and indicated the direction of coherent motion in a two-alternative forced-choice task. During the experiment, the coherency level (proportion of coherently moving dots) was fixed and determined in the practice block, ensuring that individuals achieved approximately 80% accuracy.

To model this dataset, we assumed that the drift rate remains fixed throughout the experiment. However, we allowed the threshold-setting and non-decision time to vary across speed and accuracy conditions (Since some studies showed the drift rate can be affected by speed/accuracy manipulation [e.g., Ratcliff and McKoon, 2023], we also tested free-drift rate models in which the drift rate can vary across speed/accuracy conditions, and the obtained results were similar to what we have reported here). Similar to Study 1, we fitted HCT-DDM and ECT-DDM in two NDT-informed and uninformed versions, plus an FT-DDM, which includes across-trial variability in drift rate on empirical data. The non-decision time measurements were again obtained from a standard cumulative HMP fit (as reported in Section 4.4 of Weindel et al., 2024). The HMP solution identified three events in the speed-focused condition and four events in the accuracy-focused condition. The sequence and timing of these events are presented in Figure 11. The non-decision time was taken as the difference between the RT and the duration of the second-to-last stage.

As in Study 1, quantitative results showed that an NDT-informed model is the best-performing. The estimated mean parameter values and the model comparison results are presented in Table 3. The BIC-based model comparison suggests that the best-fitting model is the NDT-informed ECT-DDM. Therefore, the main result of the first study, which was an improvement in model fit when we constrain non-decision time, was replicated in this study.

The obtained thresholds differed notably between speed and accuracy conditions, with both the starting level and the decay rate being lower in the speed condition. The estimated threshold dynamics for each participant, along with the mean group estimated threshold, are presented in Figure 12 (see Appendix 6 for the estimated collapsing threshold by uninformed models). Similar to Study 1, the estimated thresholds for individuals suggest that most participants adopt a collapsing threshold stopping rule. Also, participants in this study exhibited a lower starting threshold in the speed condition compared to the accuracy condition. However, in contrast to Study 1, the decay rate in the speed condition was lower. This reduction in the decay rate under the speed emphasis condition is due to the huge effect of speed-accuracy threshold manipulation on the starting threshold.

The NDT-informed models' predictions are depicted in Figure 13. This figure shows that the FT-DDM overestimates the last RT quantiles for both correct and incorrect responses in both speed and accuracy conditions. However, the qualitative predictions of CT-DDMs align more closely with the empirical data. It is also worth noting that all the considered computational models misfit the incorrect responses in the speed condition. This misfit is to be linked to the presence of fast errors, specifically in the speed condition. Including the starting-point variability parameter in the model enables the model to predict fast errors. However, as the aim here was not merely to fit the data with the best possible model, but to test the NDT-informed modeling framework, we did not include starting-point variability in the model.

The results from this study replicated the main findings of the first study. The model comparison results showed that NDT-informed models fit the empirical data better than the uninformed models (in this study, NDT-informed ECT-DDM was the best-fitting model). Also, comparing these results with the results of FT-DDM reported in Appendix 7 revealed that the CT-DDMs better account for empirical data than the FT-DDM. This model comparison result, along with the high mean decay rate (reported in Table 3), supports the CT-DDMs against FT-DDMs.

Discussion

Understanding whether people adjust their decision threshold over the time course of a single decision has been a central topic in cognitive science for decades. Despite the theoretical and empirical support for CT-DDMs, the main issue limiting their application in individual differences

Figure 11. The event sequence and timing of each event estimated by the HMP method for Accuracy (top) and Speed (bottom) conditions averaged over all participants (Weindel et al., 2024).

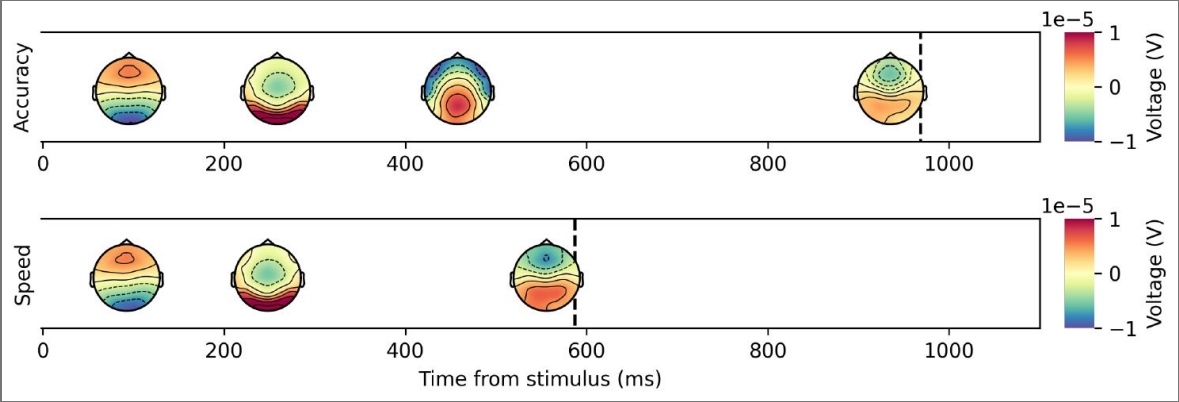


Table 3. The mean estimated parameters and goodness of fit results for Study 2.

Model	δ	τ_{SP}	τ_{AC}	θ_{SP}	θ_{AC}	λ_{SP}	λ_{AC}
NDT-informed HCT-DDM	0.51	0.27	0.31	0.76	1.99	1.31	2.29
NDT-informed ECT-DDM	0.52	0.27	0.31	0.73	1.58	0.81	0.90
NDT-informed FT-DDM	0.57	0.29	0.43	0.56	0.72	—	—
Uninformed HCT-DDM	0.51	0.28	0.28	0.77	2.27	1.06	2.58
Uninformed ECT-DDM	0.53	0.27	0.31	0.74	1.57	0.72	0.87
	$\sigma_{z,SP}$	$\sigma_{z,AC}$	BIC_{Joint}	$BIC_{Behavior}$			
NDT-informed HCT-DDM	0.35	0.56	-622.96	4164.78			
NDT-informed ECT-DDM	0.35	0.58	-669.01	4116.75			
NDT-informed FT-DDM	0.34	0.34	34.40	4466.14			
Uninformed HCT-DDM	—	—	—	4268.05			
Uninformed ECT-DDM	—	—	—	4216.85			

Figure 12. The estimated threshold dynamics from NDT-informed CT-DDMs for individuals in Study 2.

The first row illustrates the estimated hyperbolic threshold dynamics, and the second row illustrates the estimated exponential threshold dynamics. The left column shows the threshold dynamics in the speed condition, and the right column shows the accuracy condition. Each gray line corresponds to a subject. The blue line corresponds to the average group level.

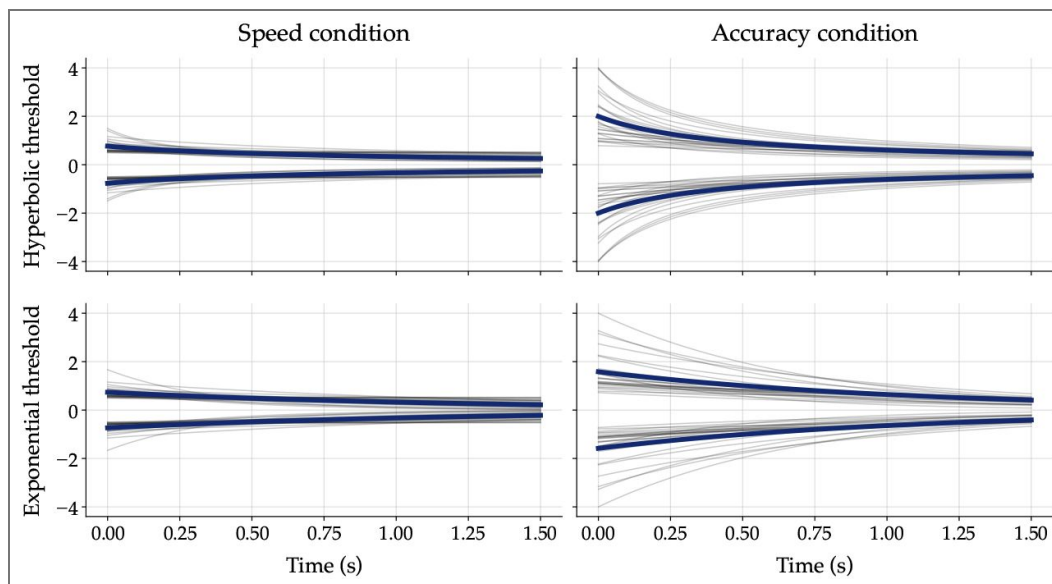
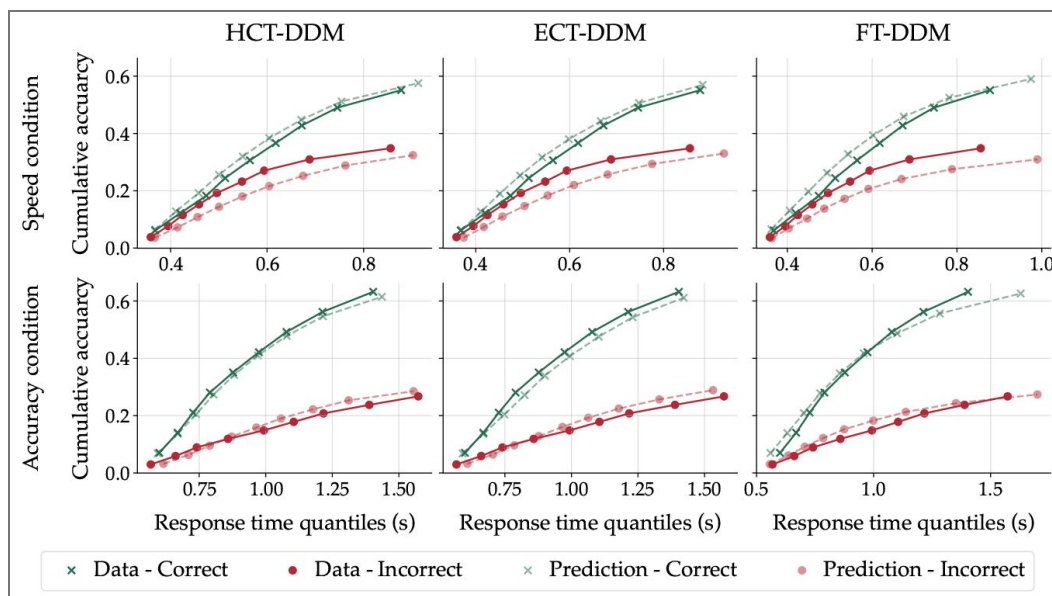


Figure 13. Prediction of NDT-informed models against empirical data for speed (top row) and accuracy (bottom row) conditions in Study 2.

In each panel, the x-axis represents the response time quantiles in seconds, and the y-axis represents the cumulative choice proportion.



studies is the poor parameter recovery of these models. The present study addressed a long-standing question in computational cognitive neuroscience by proposing a diffusion modeling approach for the reliable estimation of time-dependent parameters. Specifically, we introduced a joint modeling framework that constrains non-decision time using additional observations extracted from neural signals. Simulation results demonstrated that incorporating non-decision time information significantly enhances the parameter recovery of collapsing threshold dynamics, making previously unrecoverable parameters recoverable. To test the robustness of this finding, the simulation results were replicated using two theoretically motivated forms of threshold dynamics: hyperbolic and exponential. While the precision of parameter estimation varied between these functional forms, the improvement in estimation was consistent across both models. That is, regardless of the threshold functional form, using noisy non-decision time measurements to inform the model consistently improved the reliability of parameter estimation. Importantly, the exponential collapsing threshold exhibited substantially better parameter recovery and superior model recovery performance, suggesting that this specification may be preferable in future cognitive modeling applications. Furthermore, the simulations showed that even under conditions of high noise in the non-decision time measurements, the parameter recovery of the NDT-informed model remained superior to that of the uninformed model. Additionally, the results of the model recovery simulation study showed that the NDT-informed modeling approach can reliably be used to infer the underlying evidence accumulation process. The model recovery study revealed that incorporating the estimated decay rate into model classification substantially increases model recovery precision.

In addition to the simulation studies, we reanalyzed two datasets on perceptual decision-making to demonstrate the feasibility and applicability of the proposed method in real empirical settings. In both studies, EEG recordings were available, and we used the HMP estimates from neural signals presented in the study by Weindel et al. 2024 [\[1\]](#) as measurements of non-decision time on a trial-by-trial basis. The cognitive modeling results revealed three main findings. First, informing CT-DDMs with non-decision time measurement improved model fit compared to their uninformed counterparts. This improvement likely reflects more accurate parameter estimation enabled by the additional information. In other words, constraining the non-decision time using neural measurements led the optimizer to estimate the CT-DDM parameter more accurately and, as a result, improve the fit to empirical data. Second, across both datasets, CT-DDMs outperformed FT-DDM in terms of model fit, regardless of whether the models were informed by non-decision time (see Appendix 7). This result, along with the better qualitative predictions of CT-DDMs, supports the notion that individuals adjust their decision thresholds dynamically within a single decision. Third, the estimated mean decay rate (λ) was notably high in both datasets, especially in the NDT-informed CT-DDMs — further supporting the presence of threshold collapse, at least in the analyzed datasets. Importantly, we found the evidence in two different types of experimental paradigms. In the second study, the information available to the participant for making their decision changed over the course of a single trial. However, in the first study, the stimulus remained static throughout the trial, and the amount of information remained constant. Most studies that have found evidence for urgency have employed information-changing paradigms (Trueblood et al., 2021 [\[2\]](#); Evans et al., 2020a [\[3\]](#); Evans and Hawkins, 2019 [\[4\]](#); Khodadadi et al., 2017 [\[5\]](#); Gluth et al., 2012 [\[6\]](#), 2013 [\[7\]](#)), and the support for collapsing the threshold in experiments with static stimuli is limited (see Olschewski et al., 2025 [\[8\]](#)).

The computational modeling results contribute to the literature by providing empirical support for collapsing threshold models. The debate on whether people become more urgent as time passes is still ongoing. Although there are some empirical and theoretical supports for CT-DDMs, some studies based on quantitative model comparison across several experiments did not find strong support for collapsing threshold models (Voskuilen et al., 2016 [\[9\]](#); Smith and Ratcliff, 2022 [\[10\]](#); Milosavljevic et al., 2010 [\[11\]](#); Karşilar et al., 2014 [\[12\]](#)) or found that the collapsing threshold only happens in some specific circumstances (Hawkins et al., 2015 [\[13\]](#)). One reason for the mixed findings in the literature is the unreliability of parameter estimation in CT-DDMs (Evans et al., 2020b [\[14\]](#); Murrow and Holmes, 2024a [\[15\]](#)), along with high model mimicry in diffusion models (Khodadadi and Townsend, 2015 [\[16\]](#)), particularly when key parameters, such as the drift rate, vary

across trials. However, the simulation results showed that based on NDT-informed diffusion modeling, we can reliably estimate the model's parameters and infer the underlying evidence accumulation process. Thus, the proposed NDT-informed diffusion modeling approach can provide a stronger conclusion.

In addition to evidence supporting CT-DDMs, the computational modeling results also revealed that different types of speed-accuracy trade-off manipulations have distinct effects on the dynamics of the decision threshold. In Study 1, participants received verbal instructions emphasizing speed. This manipulation led to a faster decay rate and a decrease in the starting threshold. However, in Study 2, where participants faced a clear response deadline, the manipulation of speed caused a decrease in the decay rate. This decrease is mostly due to the large decrease in the starting threshold caused by speed emphasis in study 2. Task- and instruction-dependent effects of speed-accuracy trade-off manipulation have already been reported in several studies (e.g., Katsimpokis et al., 2020 [↗](#); Evans et al., 2019 [↗](#)), and our results are consistent with this pattern.

One promising avenue for future research is the investigation of individual differences in the shape of threshold dynamics. In the present work, we examined two theoretically motivated forms — hyperbolic and exponential — and fitted them at the individual level. However, recent evidence suggests that threshold dynamics may vary substantially across individuals (e.g., Kira et al., 2025 [↗](#)). Non-parametric approaches have been proposed to recover the full threshold trajectory without assuming a specific functional form (e.g., Kira et al., 2025 [↗](#); Fudenberg et al., 2020 [↗](#)). Findings from Kira et al. 2025 [↗](#) indicate marked heterogeneity across participants. Understanding such variability could provide critical insights into why choice behavior sometimes deviates from optimality. A limitation of current non-parametric approaches, however, is that they require a vast number of trials per participant, making them impractical for studies with larger sample sizes. Combining NDT-informed diffusion modeling with non-parametric methods may help alleviate this limitation by reducing the trial requirements, thereby making individual differences studies on threshold dynamics more accessible to empirical investigation.

Limitation

The collapsing threshold dynamics considered in this work (e.g., exponential and hyperbolic) impose a monotonically decreasing threshold over time. Although these dynamics are theoretically well motivated (Fudenberg et al., 2018 [↗](#); Frazier and Yu, 2007 [↗](#)) and have been employed in several previous studies (e.g., Olschewski et al., 2025 [↗](#); Milosavljevic et al., 2010 [↗](#); Voskuilen et al., 2016 [↗](#)), some research has proposed delayed-collapsing threshold models (e.g., Diederich and Oswald, 2016 [↗](#)), in which the onset of threshold collapse does not coincide with the onset of evidence accumulation. Estimating such delayed-collapsing models may require more than simply constraining non-decision time, as the onset of threshold collapse must also be identified. A promising approach for addressing this challenge is the HMP method, which may provide additional temporal information about distinct cognitive processing stages. In particular, HMP may allow the onset of threshold collapse to be estimated as a separate cognitive stage. Future research should therefore investigate the estimation of multi-stage evidence accumulation models (e.g., Diederich and Oswald, 2016 [↗](#); Diederich and Colonius, 2021 [↗](#)) within the HMP framework.

Distributional assumption on non-decision time measurement

Although we assumed that non-decision time measurements follow a log-normal distribution, alternative distributional assumptions are also plausible. In particular, some studies have modeled non-decision time using symmetric distributions such as the normal distribution (e.g., Ghaderi-Kangavari et al., 2023 [↗](#); Kira et al., 2025 [↗](#)). In Appendix 8, we formalize the joint NDT-informed framework under the assumption that non-decision time measurements are normally distributed and report the corresponding parameter recovery results. The findings show that the improvement in recovering collapsing threshold parameters persists under this alternative

distributional assumption. Although minor quantitative differences emerge across distributional specifications, the enhanced reliability of threshold parameter estimation is not confined to a specific assumption and generalizes across different distributional forms.

Would NDT-informed modeling lead to a different conclusion?

It is important to note that the NDT-informed diffusion modeling approach can sometimes yield conclusions that differ from those obtained using traditional modeling methods. For example, in Study 1, the results of the uninformed ECT-DDM suggest that the speed–accuracy manipulation does not affect the starting threshold. In contrast, both the NDT-informed ECT-DDM and NDT-informed HCT-DDM indicate that the speed–accuracy trade-off manipulation does influence the starting threshold (see Table 2 [↗](#)). This discrepancy arises from unreliable parameter estimation in the uninformed CT-DDMs. However, the simulation results and sensitivity analyses reported in this paper demonstrate that the parameter estimates provided by the NDT-informed diffusion modeling framework are substantially more reliable. Therefore, any parameter-based inference concerning the underlying mechanisms of decision-making, the effects of experimental manipulations, or individual and group differences should be drawn using the NDT-informed diffusion modeling framework.

Generalization to other computational models

Although the primary focus of the present study was on two-alternative choice tasks, understanding urgency is also crucial in the context of multi-alternative and multi-attribute decisions (e.g., Tajima et al., 2019 [↗](#); Gluth et al., 2026 [↗](#); Busemeyer et al., 2019 [↗](#)), as well as in decisions involving a continuous range of options (e.g., circles, lines, or arcs; Hadian Rasanan et al., 2025 [↗](#), 2026 [↗](#), 2024a [↗](#)). To examine the generalizability of the proposed method, we conducted a parallel parameter recovery study using the n -dimensional hyper-spherical diffusion model (Smith and Corbett, 2019 [↗](#); Smith, 2016 [↗](#); Hadian Rasanan et al., 2026 [↗](#)). This model is suitable for modeling multi-alternative/multi-attribute choices (Kvam, 2019a [↗](#); Smith and Corbett, 2019 [↗](#); Smith, 2019 [↗](#)) as well as decision-making tasks with continuous responses (Kvam, 2019b [↗](#); Smith et al., 2023 [↗](#), 2020 [↗](#); Hadian Rasanan et al., 2026 [↗](#)). The results of this simulation study are presented in Appendix 9. Consistent with the findings from the one-dimensional DDM, the results show that informing the n -dimensional diffusion model with non-decision time measurements significantly improves the precision of threshold dynamics estimation (see Appendix 9). These findings suggest that the improvement in parameter recovery of the time-dependent threshold is not limited to binary choice tasks but also extends to the more general n -dimensional hyper-spherical diffusion model. This makes the proposed method applicable to a broader range of decision-making paradigms involving multi-alternative, multi-attribute, or continuous response decision tasks.

The proposed NDT-informed diffusion modeling approach can also be applied to parameter estimation in urgency gating models. Unlike CT-DDMs, urgency-gating models assume a fixed decision threshold, but allow the incoming evidence to scale as time progresses (Trueblood et al., 2021 [↗](#); Cisek et al., 2009 [↗](#)). However, similar to CT-DDMs, urgency-gating models also suffer from poor parameter recovery, as reported by Evans et al. 2020b [↗](#). Thus, the issue of parameter unreliability extends to these models. Recently, Smith and Ratcliff, 2022 [↗](#) formalized the connection between CT-DDMs and urgency gating models, demonstrating that the multiplicative urgency gating model is mathematically equivalent to a hyperbolic collapsing threshold diffusion model. In other words, urgency gating models and hyperbolic CT-DDMs yield identical behavioral predictions, and their parameters can be transformed into one another. Consequently, the findings from our simulation studies on hyperbolic collapsing threshold dynamics also apply to urgency gating models (Trueblood et al., 2021 [↗](#); Cisek et al., 2009 [↗](#)).

Another way to implement urgency in the decision process is to consider an independent accumulator for timing. The time-based racing diffusion model (Hawkins and Heathcote, 2021 [↗](#)) considers independent accumulators corresponding to evidence for available options, as well as an additional time-based accumulator for timing. This time-based accumulator behaves like an

inner clock, and when it exceeds the threshold before the other evidence accumulators, an option will be randomly selected. Similar to the CT-DDM, this model can also generate Gaussian-like distributions of RTs and shows a good fit on empirical data. This model can also predict faster error responses than the correct response, the pattern that is observed in the speed condition of Study 2. However, the authors reported poor parameter recovery for the onset of the timing process (Hawkins and Heathcote, 2021 [↗](#)). Thus, the reliability issue here specifically concerns the estimation of the shift parameter of the timing accumulator. Informing the model with external estimates of non-decision time might, therefore, improve parameter recovery in this model as well.

More generally, collapsing thresholds are not the only evidence accumulation mechanisms that suffer from reliability issues. Poor parameter recovery has also been reported for time-dependent drift rate models, such as diffusion models of conflict or diffusion models with leakage (White et al., 2018 [↗](#); Evans et al., 2020b [↗](#)). Although several studies have attempted to improve parameter recovery in these models (e.g., Hübner and Pelzer, 2020 [↗](#); Miletic et al., 2017 [↗](#)), the problem remains unresolved. A promising direction for enhancing parameter estimation in such models is to constrain the drift rate dynamics using neural data. In particular, several studies have shown that Centro-parietal positivity (CPP) in EEG signals is closely linked to the process of evidence accumulation (e.g., O'Connell et al., 2012 [↗](#); Steinemann et al., 2018 [↗](#); Kohl et al., 2020 [↗](#); Kelly et al., 2021 [↗](#); Grogan et al., 2025 [↗](#)). Furthermore, a recent application of HMP has shown that single-trial estimates of this potential can be extracted with reasonable accuracy (Weindel et al., 2025 [↗](#)). Thus, informing time-dependent drift rate models with CPP dynamics may improve the identifiability and reliability of their parameter estimates.

Does only non-decision time improve parameter estimation?

In this work, we focused on constraining the non-decision time and provided both theoretical justification and simulation evidence demonstrating that constraining non-decision time in the diffusion model can enhance the reliability of parameter estimation in CT-DDMs. However, it remains an open question whether constraining other parameters — such as the drift rate — can similarly improve parameter estimation in CT-DDMs. To address this, it is essential to note that, from a mathematical standpoint, assuming a fixed diffusion coefficient, there is no trade-off between drift rate and threshold parameters (Nunez et al., 2025 [↗](#); Ratcliff, 1978 [↗](#)). Therefore, constraining the drift rate using neural signals would not improve the estimation of collapsing threshold parameters.

Another neurophysiological signal that may aid in improving parameter estimation of CT-DDMs is pupil dilation. Several studies have demonstrated a correlation between pupil dilation and decision threshold (e.g., Murphy et al., 2016 [↗](#); Cavanagh et al., 2014 [↗](#)), suggesting that the pupil dilation time series may contain valuable information about threshold dynamics. However, this approach presents certain challenges. Notably, accurately identifying the time window corresponding to the decision process using only eye-tracking data is difficult. To overcome this limitation, additional sources of information — such as EEG signals — may be necessary. As a result, directly constraining threshold dynamics is more complex than constraining non-decision time. Nevertheless, the potential advantages of incorporating pupil dilation into joint modeling approaches warrant further investigation in future research.

Bias in non-decision time measurement

One of the main assumptions of NDT-informed diffusion modeling is that the true non-decision time equals the mean of the non-decision time measurement distribution. However, this assumption may not always hold in certain situations. In other words, there might be a systematic bias (i.e., overestimation or underestimation) in the non-decision time measurements. To investigate the effect of such biases in non-decision time measurements on parameter estimation, we conducted a model misspecification study in which the true non-decision time parameter deviates from the mean of non-decision time measurements (see Appendix 5). The results of the model misspecification study showed that when we have an overestimation bias in non-decision

time measurements, we observe an underestimation in the starting threshold and decay rate parameters (with high correlation with the true parameters). In contrast, when there is an underestimation bias in non-decision time measurements, we observe an overestimation in starting threshold and decay rate parameters.

Behavioral methods for estimating non-decision time

Although the focus of the current study was on extracting non-decision time from neural signal data using the HMP method, it is also worth noting that some alternative approaches do not require neural data. Historically, disentangling decision time and non-decision time has been a long-standing question in mathematical psychology and psychophysics (e.g., [Green and Luce, 1971](#) [↗](#); [Kohfeld et al., 1981](#) [↗](#); [Burbeck and Luce, 1982](#) [↗](#); [Smith, 1990](#) [↗](#); [Sheu and Ratcliff, 1995](#) [↗](#); [Verdonck and Tuerlinckx, 2016](#) [↗](#); [Bompas et al., 2025](#) [↗](#)). These attempts resulted in the development of convolutional-based methods. The core idea behind these methods is that the response time distribution can be represented by the convolution of the decision time distribution predicted by the diffusion model and the non-decision time distribution ([Smith, 1990](#) [↗](#); [Verdonck and Tuerlinckx, 2016](#) [↗](#)). Therefore, one way of using this relation is to deconvolve the non-decision time distribution from the observed RT distribution, resulting in the decision time distribution ([Smith, 1990](#) [↗](#)). Thus, running a simple one-choice reaction task (i.e., responding as soon as the stimulus is presented on the screen) with identical stimuli, which are also presented in the main task, provides an estimation of the non-decision distribution. In other words, response time in such one-choice reaction tasks includes only the perceptual encoding and motor execution, excluding the evidence accumulation process, and approximates the non-decision time in the main task. Deconvolution of the non-decision time distribution obtained from the one-choice reaction task to the response time distribution of the main task provides an approximation of decision time in the main task. This approach requires running an additional task and also has some practical issues ([Sheu and Ratcliff, 1995](#) [↗](#)).

To address these issues, [Verdonck and Tuerlinckx, 2016](#) [↗](#) proposed an alternative convolution-based approach that can estimate the distribution of non-decision time entirely based on observed data in a single experiment. They estimate the non-decision time distribution by convolving the observed RT distribution with the predicted distribution from the diffusion model in different conditions and minimizing the distance between the resulting distributions. Their results showed that their proposed method can accurately estimate the distribution of non-decision times and cancel out the effect of non-decision time on RT. Recently, [Kira et al. 2025](#) [↗](#) have employed a similar approach to cancel out the impact of non-decision time on RT and then estimate the entire threshold dynamics using a non-parametric approach.

Conclusion

The question of whether people become more urgent as time passes is one of the long-standing questions in cognitive psychology. In this work, we introduced an NDT-informed diffusion modeling framework that significantly improves the reliability of parameter estimation in CT-DDM. Therefore, this approach enables researchers to use parameter values in their inference about the underlying cognitive mechanism of the decision process and to use CT-DDM as a measurement tool to investigate individual differences. Based on the proposed NDT-informed diffusion modeling framework, we leveraged EEG estimates of non-decision time and reanalyzed two empirical datasets. We found evidence for CT-DDMs across two studies, which implies that people become more urgent as time passes.

Data availability

All codes and data supporting this paper are publicly available on:
https://github.com/AmirHoseinHadian/CTDM_recovery [↗](#)

Acknowledgements

The authors would like to thank Nathan J. Evans, Sebastian Gluth, and Amin Ghaderi-Kangavari for their insightful comments on the early version of this work.

Additional information

Funding

AHHR and JR are supported by the Swiss National Science Foundation (Grant No. 214099). GW is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (Grant No. 101066503).

Funding

Funder	Grant reference number	Author
Swiss National Science Foundation (SNF)	214099	Jörg Rieskamp Amir Hosein Hadian Rasanan
European Union Horizon	https://doi.org/10.3030/101066503	Gabriel Weindel

Author ORCID iDs

Amir Hosein Hadian Rasanan:  <https://orcid.org/0000-0001-7785-1514>

Gabriel Weindel:  <https://orcid.org/0000-0002-7592-1686>

Jörg Rieskamp:  <https://orcid.org/0000-0003-2632-8015>

Appendix 1

Trade-off between non-decision time and collapsing threshold recovery

To illustrate how imprecise estimation of non-decision time affects the recovery of collapsing threshold parameters (i.e., starting threshold and decay rate), we conducted an additional simulation study. In this study, data was generated from an uninformed CT-DDM, after which model parameters were estimated under the assumption that non-decision time was fixed rather than estimated freely. To show how imprecise estimation of non-decision time can distort threshold estimation, the non-decision time was either held fixed or contaminated with noise. Using this scenario and modulating the noise level, we can understand how and on what scale noisy estimation of non-decision time affects threshold estimation.

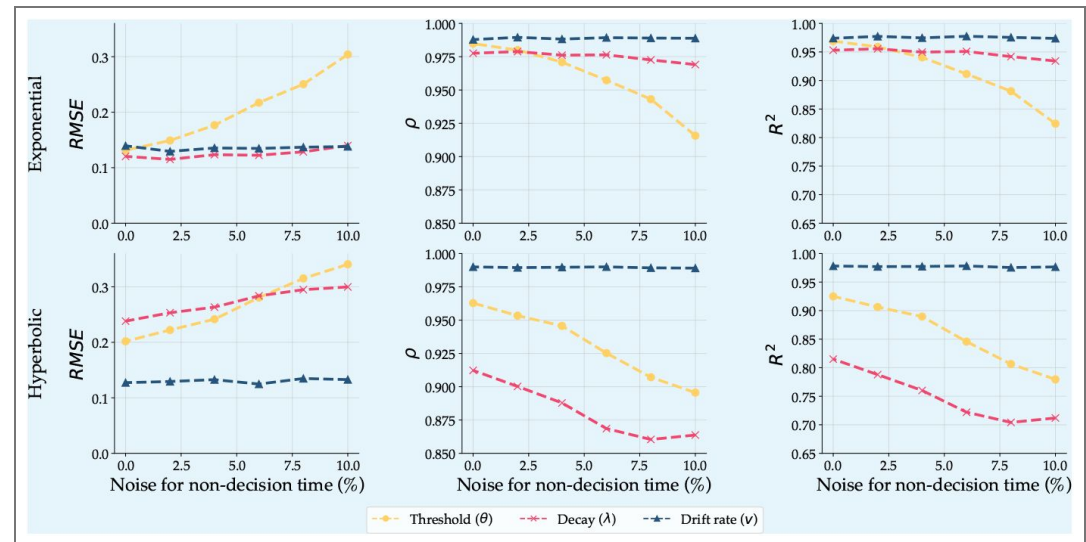
To simulate data, first, we generated sample parameters from the following uniform distributions:

$$\begin{aligned}\theta &\sim \mathcal{U}[1.5, 4], \\ \lambda &\sim \mathcal{U}[0.1, 2], \\ \delta &\sim \mathcal{U}[0, 3], \\ \tau &\sim \mathcal{U}[0.05, 1].\end{aligned}$$

After sampling the parameters, we generated response time and choice for 500 trials using the drawn parameters. Then, we estimated parameters using the simulated data. But during the estimation procedure, the non-decision time parameter was treated as known and held constant. In other words, we only estimated θ , λ , and δ . The fixed non-decision time value was either exact or contaminated with uniform noise at five levels (2%, 4%, 6%, 8%, or 10%). The contaminated non-decision time is obtained by $\tau_{\text{contaminated}} = (1 + \varepsilon) \times \tau$, where $\varepsilon \sim \mathcal{U}[-n, +n]$, and n is either 0.02, 0.04, 0.06, or 0.1.

Figure 1  illustrates the effect of noisy non-decision time estimation on the recovery of other model parameters in both exponential and hyperbolic CT-DDMs. As shown in this figure, increasing noise in the non-decision time estimates substantially degrades the estimation of the

starting threshold (θ) and decay rate (λ) parameters. Specifically, the root mean square error (RMSE) for these parameters increases as the noise level in non-decision time increases. In addition, both the correlation coefficient (ρ) and the R^2 value for these parameters decrease with increasing non-decision time noise. In contrast, these metrics for the drift rate remain largely unchanged, indicating that drift rate estimation is relatively unaffected by imprecise non-decision time estimates. These simulation results are consistent with the theoretical arguments presented in the main text and confirm the existence of a trade-off between non-decision time and collapsing threshold parameters.



Appendix 1—figure 1. Illustration of the imprecise estimation of non-decision time on other parameters of the CT-DDM.

Appendix 2

Model fitting using the integral equation method

Since the diffusion models with time-dependent thresholds do not have an exact likelihood function, we need to approximate it using a numerical method. The integral equation method provides a reliable numerical procedure for approximating the first-passage time distribution of diffusion models with time-dependent parameters (Smith, 2000; Smith and Ratcliff, 2022). This method was first introduced by Buonocore et al. (1987, 1990) and then employed in psychological research frequently (e.g., Smith, 2000; Voskuilen et al., 2016; Smith and Ratcliff, 2022; Evans et al., 2020a; Hadian Rasanan et al., 2025; Smith, 2023; Zhang et al., 2014). In this method, the first-passage time distributions of upper and lower boundaries are obtained by solving the following system of linear integral equations (in mathematics, this equation is known as the Volterra integral equation of the second kind, Wazwaz, 2011):

$$\begin{aligned} g_u[b_u(t), t | z, 0] = & -2\Psi[b_u(t), t | z, 0] \\ & + 2 \int_0^t g_u[b_u(\tau), \tau | z, 0] \Psi[b_u(t), t | b_u(\tau), \tau] d\tau \\ & + 2 \int_0^t g_l[b_l(\tau), \tau | z, 0] \Psi[b_u(t), t | b_l(\tau), \tau] d\tau \end{aligned} \quad (7)$$

and

$$\begin{aligned}
 g_l[b_l(t), t | z, 0] = & 2\Psi[b_l(t), t | z, 0] \\
 & - 2 \int_0^t g_u[b_u(\tau), \tau | z, 0] \Psi[b_l(t), t | b_u(\tau), \tau] d\tau \\
 & - 2 \int_0^t g_l[b_l(\tau), \tau | z, 0] \Psi[b_l(t), t | b_l(\tau), \tau] d\tau
 \end{aligned} \tag{8}$$

where $g_u[b_u(t), t | z, 0]$ and $g_l[b_l(t), t | z, 0]$ represent the first-passage time distributions of crossing the upper and lower time-dependent thresholds, respectively. Also, Ψ is the kernel function and is defined as follows (Giorno et al., 1989; Richter et al., 2023):

$$\Psi[b(t), t | y, \tau] = \frac{f[b(t), t | y, \tau]}{2} \left(b'(t) - \frac{b(t) - y}{t - \tau} \right). \tag{9}$$

In this equation, $f[x, t | y, \tau]$ represents the free transition density function. For the one-dimensional diffusion process with drift rate δ , the free transition density function is defined as follows (Giorno et al., 1989; Richter et al., 2023):

$$f[x, t | y, \tau] = \frac{1}{\sqrt{2\pi\sigma^2(t - \tau)}} \exp\left(-\frac{(x - y - \delta(t - \tau))^2}{2\sigma^2(t - \tau)}\right). \tag{10}$$

To approximate the solution of this integral equation, various numerical methods are available in the literature (see Chapter 3 in Wazwaz, 2011). However, the left-point rectangular scheme is the most straightforward and efficient numerical approximation scheme, offering relatively high precision in likelihood approximation (Richter et al., 2023). So, by discretizing the temporal space using N equidistant points (i.e., $\Delta t = \frac{T_{\max}}{N}$, we considered $T_{\max}$ equal to the longest RT) the approximation of the first-passage time at the first time step is obtained as follows (Buonocore et al., 1990; Smith, 2000):

$$\begin{aligned}
 \tilde{g}_u[b_u(\Delta t), \Delta t | z, 0] = & -2\Psi[b_u(\Delta t), \Delta t | z, 0] \\
 \tilde{g}_l[b_l(\Delta t), \Delta t | z, 0] = & 2\Psi[b_l(\Delta t), \Delta t | z, 0].
 \end{aligned} \tag{11}$$

Also, for the next time steps (i.e., $t_i = i\Delta t$ for $i = 2, 3, \dots$) we can use the following approximation scheme (Buonocore et al., 1990; Smith, 2000):

$$\begin{aligned}
 \tilde{g}_u[b_u(i\Delta t), i\Delta t | z, 0] = & -2\Psi[b_u(i\Delta t), i\Delta t | z, 0] \\
 & + 2\Delta t \sum_{j=1}^{i-1} \tilde{g}_u[b_u(i\Delta t), i\Delta t | z, 0] \Psi \\
 & [b_u(i\Delta t), i\Delta t | b_u(j\Delta t), j\Delta t] \\
 & + 2\Delta t \sum_{j=1}^{i-1} \tilde{g}_l[b_l(i\Delta t), i\Delta t | z, 0] \Psi \\
 & [b_u(i\Delta t), i\Delta t | b_l(j\Delta t), j\Delta t]
 \end{aligned} \tag{12}$$

$$\begin{aligned}
 \tilde{g}_l[b_l(i\Delta t), i\Delta t | z, 0] = & 2\Psi[b_u(i\Delta t), i\Delta t | z, 0] \\
 & - 2\Delta t \sum_{j=1}^{i-1} \tilde{g}_u[b_u(i\Delta t), i\Delta t | z, 0] \Psi \\
 & [b_l(i\Delta t), i\Delta t | b_u(j\Delta t), j\Delta t] \\
 & - 2\Delta t \sum_{j=1}^{i-1} \tilde{g}_l[b_l(i\Delta t), i\Delta t | z, 0] \Psi \\
 & [b_l(i\Delta t), i\Delta t | b_l(j\Delta t), j\Delta t]
 \end{aligned} \tag{13}$$

After approximating the first-passage time distributions corresponding to the upper and lower thresholds, the negative log-likelihood can be estimated using:

$$-LL = -\sum_{i=1}^{N_u} \log(\tilde{g}_u[b_u(RT_i^u - \tau), RT_i^u - \tau | 0, 0]) \\ - \sum_{i=1}^{N_l} \log(\tilde{g}_l[b_l(RT_i^l - \tau), RT_i^l - \tau | 0, 0])$$

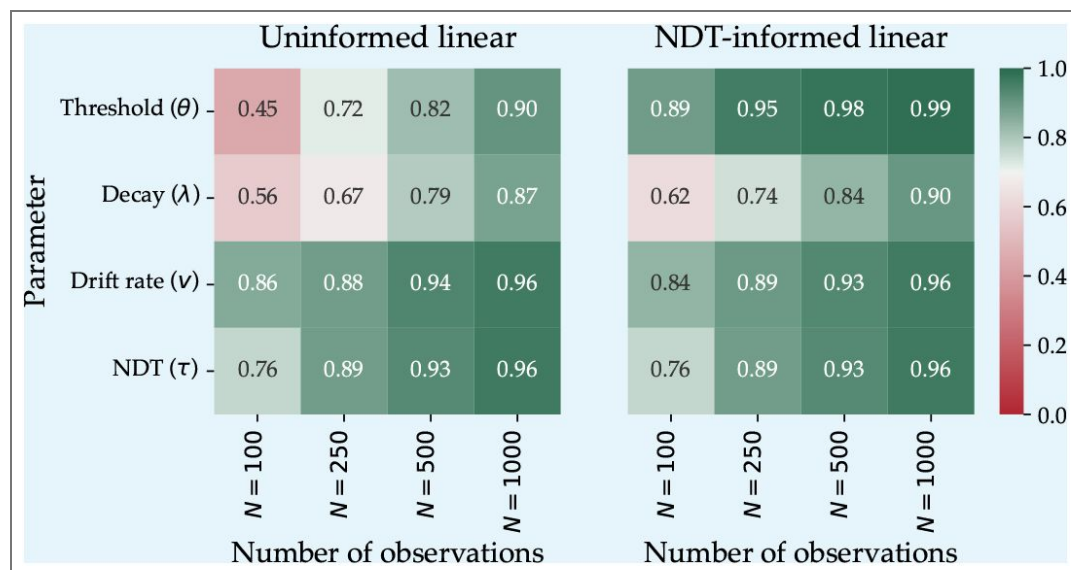
where RT_i^u and RT_i^l represent the response times ended at the upper threshold and lower threshold, respectively. Moreover, N_u is the number of trials in which the process stopped at the upper threshold, and similarly, N_l is the number of trials in which the process terminated at the lower threshold.

Appendix 3

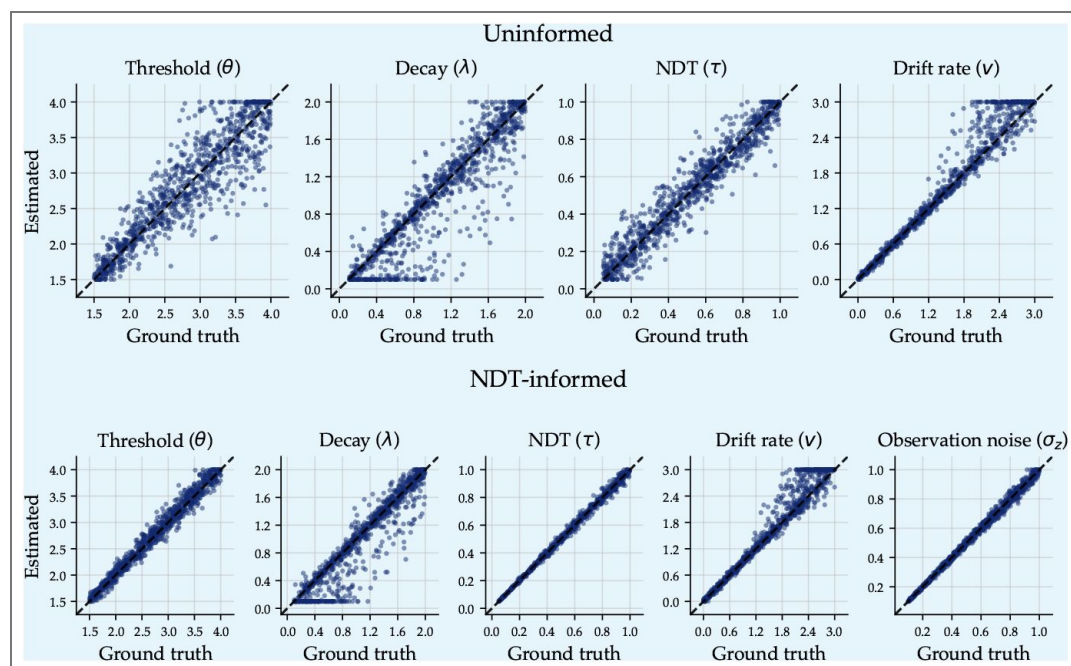
Simulation results for linear collapsing threshold diffusion model

Although there are strong theoretical motivations for considering nonlinear collapsing threshold dynamics in diffusion models, linear CT-DDMs are frequently used in the literature. This widespread use is largely due to the relatively better parameter recovery of linear CT-DDMs compared to their nonlinear counterparts. As discussed earlier, [Evans et al., 2020b](#) conducted a systematic assessment of parameter recovery in CT-DDMs and found that the parameters of linear CT-DDMs can be reliably identified given a sufficiently large number of trials. To examine whether constraining non-decision time can further improve parameter recovery and reduce the number of trials required for reliable estimation, we conducted a simulation study comparing NDT-informed and uninformed linear CT-DDMs. In this study, we assumed a linearly collapsing decision threshold (i.e., $b_u(t) = -b_l(t) = \theta - \lambda \times t$). All other aspects of the simulation design were identical to those used in the simulation study reported in the main text.

[Figure 1](#) shows the R^2 values for the parameters of both uninformed and NDT-informed linear CT-DDMs across different numbers of trials. In addition, [Figure 2](#) displays the relationship between the estimated parameters and the true generating parameters for the simulation with 500 trials. Consistent with the simulation results reported by [Evans et al., 2020b](#), the results for the uninformed linear CT-DDM indicate that the parameters become identifiable when a sufficiently large number of trials is available (i.e., more than 500 trials). However, as with nonlinear CT-DDMs, constraining the non-decision time substantially improves the estimation of collapsing-threshold parameters. Notably, the starting threshold parameter can be reliably identified with as few as 100 trials. Thus, although the parameters of the linear CT-DDM can be accurately estimated given enough data, constraining non-decision time markedly reduces the number of trials required for precise parameter estimation.



Appendix 3—figure 1. R^2 values measuring the agreement between estimated and ground truth parameters in uninformed (left) and NDT-informed (right) linear collapsing threshold diffusion models as a function of the number of observations.



Appendix 3—figure 2. Estimated vs. true parameter values for two modeling approaches. The top row shows parameter recovery for the linear collapsing threshold diffusion model estimated without additional constraints (“Uninformed”), while the bottom row shows recovery for the joint model that incorporates additional non-decision time observations (“NDT informed”). Each point represents a recovered parameter estimate plotted against its corresponding true data-generating value. Dashed diagonal lines indicate perfect recovery.

Appendix 4

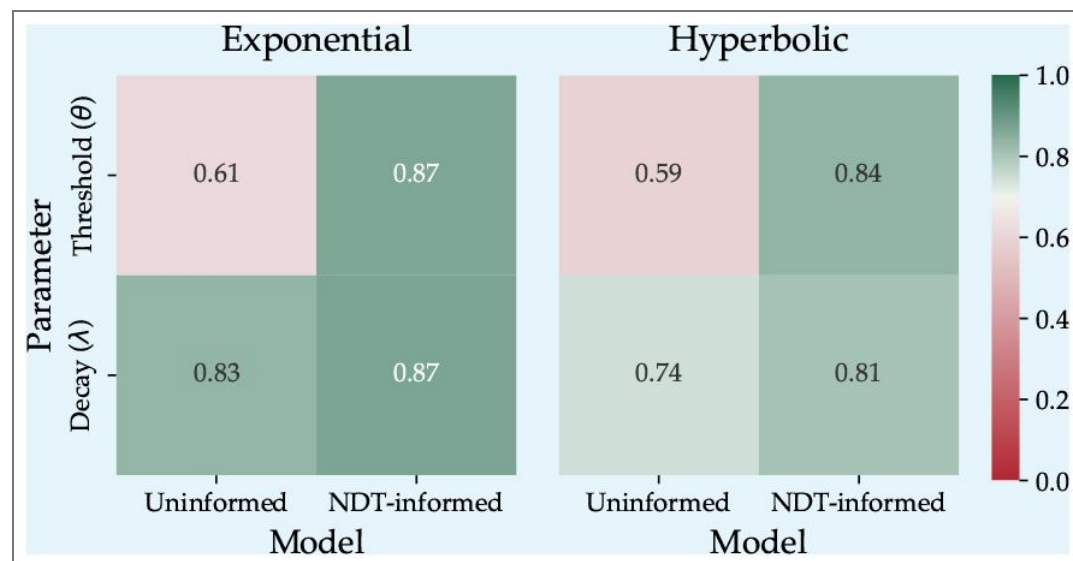
Sensitivity to estimation method

Sometimes, problems in parameter recovery are specific to a particular estimation method and may not occur with others. To ensure the robustness of our results, we conducted the same simulation study using an alternative estimation approach – Amortized Bayesian Inference (ABI;

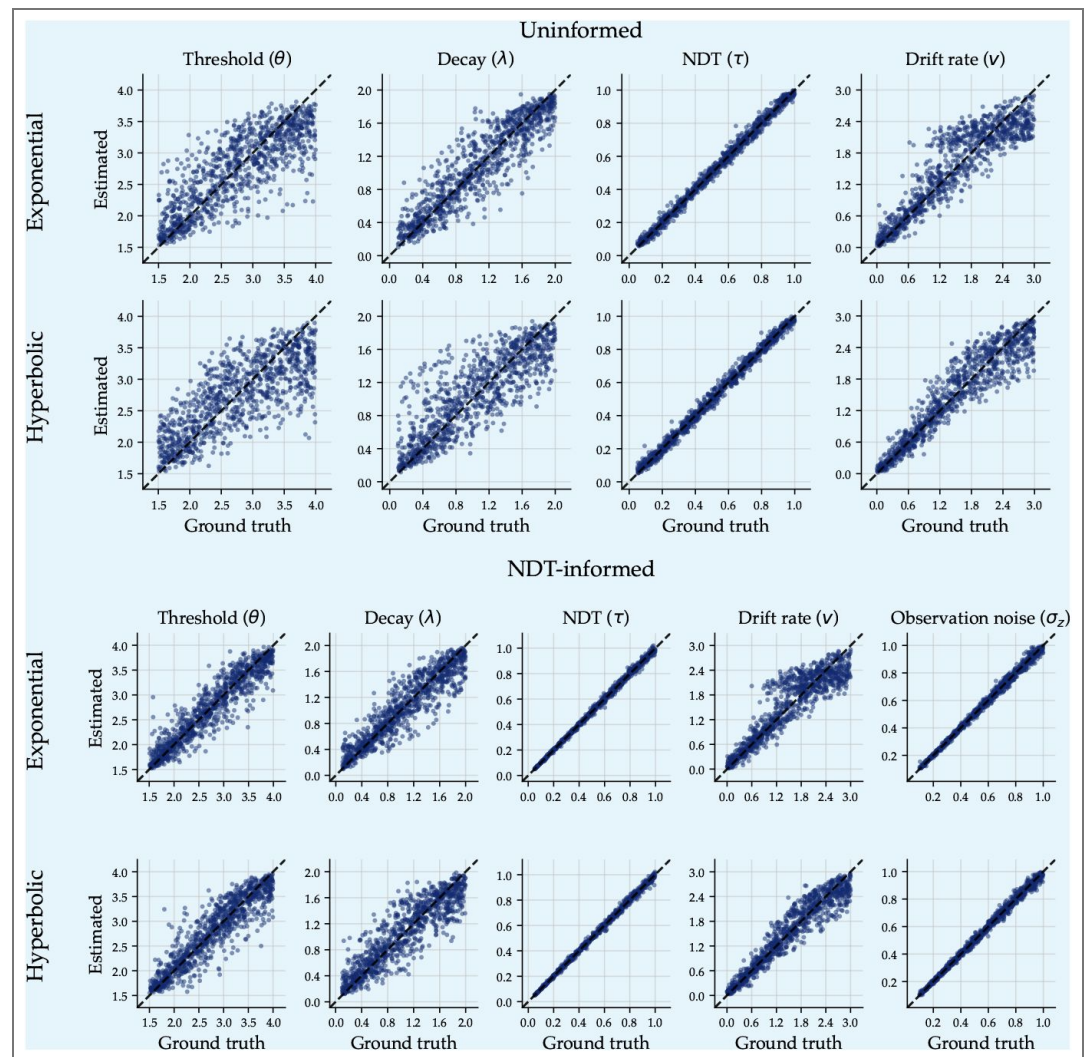
Radev et al., 2020 [\[1\]](#); Bürkner et al., 2023 [\[2\]](#)). ABI is a modern inference technique that leverages neural networks to approximate posterior distributions efficiently. Instead of performing Bayesian inference from scratch for each new dataset, ABI “amortizes” the computational cost by learning a mapping from observed data to posterior distributions across many simulated datasets.

For this purpose, we utilized the BayesFlow Python library (Radev et al., 2023b [\[3\]](#)), which provides neural network architectures tailored for ABI. Our inference pipeline consists of two main components: a set transformer that encodes the trial-wise data into maximally informative summary statistics, and an invertible neural network (the inference network) that takes these summaries — along with the generative parameters — and learns to approximate the posterior distribution. We trained four separate neural approximators, corresponding to each of the models examined: NDT-informed ECT-DDM, NDT-informed HCTDDM, Uninformed ECT-DDM, and Uninformed HCT-DDM. Each network was trained on 2 400 000 simulated datasets, with each dataset containing 500 trials and parameters sampled from prior distributions that matched those used in the main simulation study. Further implementation details, including network architectures and hyperparameters, are available in our publicly accessible GitHub repository: (https://github.com/AmirHoseinHadian/CTDM_recovery [\[4\]](#))

Figure 1 [\[5\]](#) displays the R^2 values for the threshold parameters (i.e., θ and λ) under both the no-constraint and joint models. Applying constraints on non-decision time significantly enhances parameter recovery accuracy, with the greatest improvement observed in the Hyperbolic collapsing threshold model. Compared to the integral equation estimation method, parameter recovery for the non-informed models was significantly better – again, especially for the Hyperbolic collapsing boundary. In contrast, R^2 values for the NDT-informed parameters were comparable across both methods. For completeness, Figure 2 [\[6\]](#) presents the full parameter recovery results for all four models. This figure illustrates the estimated parameters against the true generating parameters for each model.



Appendix 4—figure 1. R^2 values measuring the agreement between estimated and ground truth threshold parameters in exponential (left) and hyperbolic (right) collapsing boundary models.



Appendix 4—figure 2. Estimated vs. true parameter values for two modeling approaches. The top two rows show parameter recovery for models estimated without additional constraints (“Uninformed”), while the bottom two rows show recovery for joint models that incorporate additional non-decision time observations (“NDT informed”). Each point represents a recovered parameter estimate plotted against its corresponding true data-generating value. Rows correspond to different functional forms (Exponential and Hyperbolic). Dashed diagonal lines indicate perfect recovery.

Appendix 5

Bias in non-decision time measurements

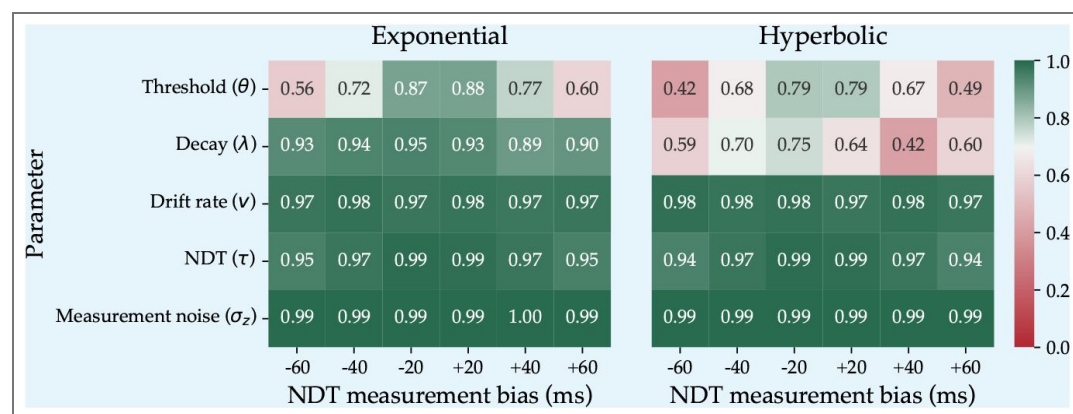
To investigate the impact of bias in non-decision time measurement on parameter estimation of NDT-informed CT-DDMs, we conducted a simulation study. The simulation setup for data generation was identical to the one considered in the main text, except that the mean of non-decision time measurements (Z_n) distribution deviated from the true non-decision time parameter (τ). To generate non-decision time measurements, we considered the following biased distribution:

$$\log(Z_n) \sim \mathcal{N}\left(\log(\tau + \varepsilon) - \frac{\sigma_z^2}{2}, \sigma^2\right)$$

where ε represents the bias in the scale of milliseconds. We considered six levels of bias in measurements (i.e., -60, -40, -20, +20, +40, and +60 milliseconds). These values cover systematic bias in the range of 60 milliseconds underestimation to 60 milliseconds

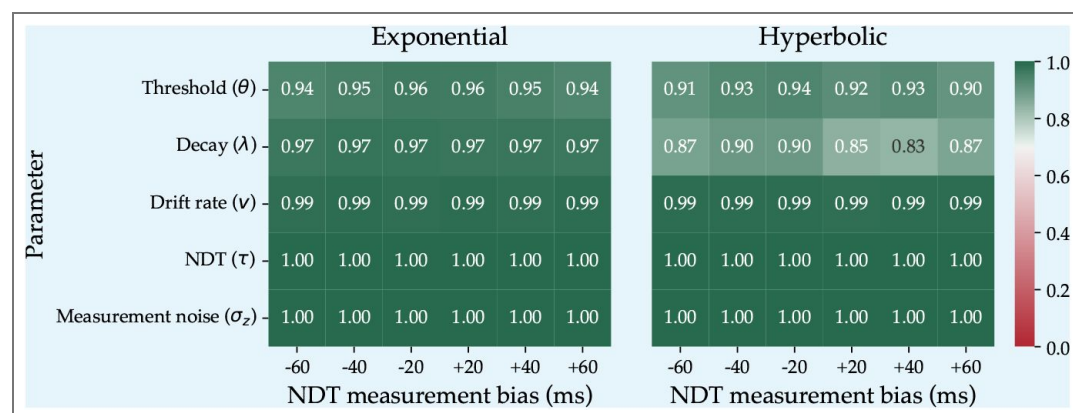
overestimation. This deviation is not the maximum deviation; rather, it represents a systematic bias in addition to the variability in the non-decision time measurements.

Figure 1 illustrates the effect of measurement bias on parameter estimation based on the R^2 value. The results suggest that the threshold parameters are sensitive to measurement bias. This sensitivity is smaller for the exponential model. At first glance, this figure may suggest that the recovery of threshold parameters becomes increasingly unreliable as the bias increases. However, it is worth noting that these low R^2 values result from bias in parameter estimation when there is a bias in non-decision time measurements.



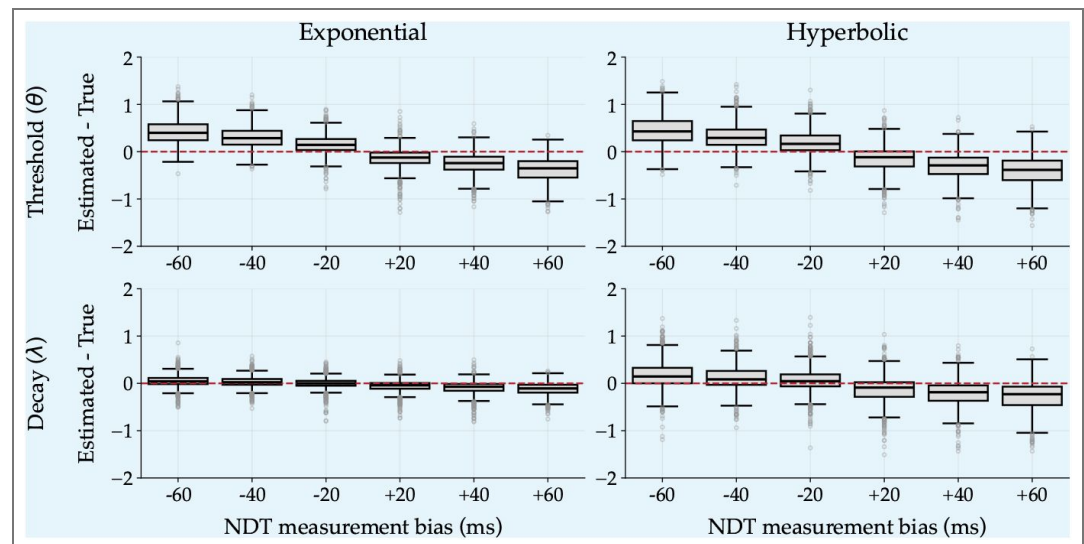
Appendix 5—figure 1. R^2 values measuring the agreement between estimated and ground truth parameters in exponential (left) and hyperbolic (right) collapsing threshold diffusion models as a function of bias in non-decision time measurements.

To illustrate this, Figure 2 shows the correlation between the true parameters and the estimated parameters in the presence of bias in non-decision time measurement. The obtained correlation values are very high, suggesting that the poor R^2 values reported in Figure 1 are mostly due to a bias in parameter estimation rather than variability or unreliability in the parameter estimation.



Appendix 5—figure 2. The correlation values measuring the agreement between estimated and ground truth parameters in exponential (left) and hyperbolic (right) collapsing threshold diffusion models as a function of bias in non-decision time measurements.

Figure 3 presents the effect of bias in non-decision time measurements on threshold parameters. Underestimation in non-decision time leads to overestimation in the starting threshold and decay rate. Conversely, overestimation in non-decision time leads to underestimation in both the starting threshold and non-decision time. The effect of bias is larger on the starting threshold than on the decay rate. Moreover, the exponential model was less sensitive to the bias in non-decision time measurements than the hyperbolic model.

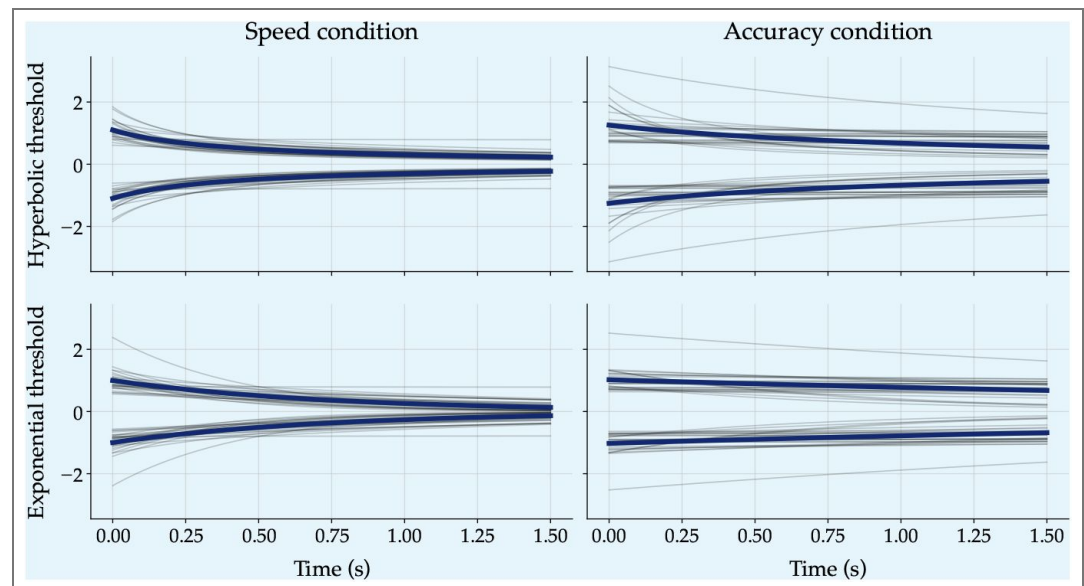


Appendix 5—figure 3. The effect of bias in non-decision time measurements on starting threshold (upper panels) and decay rate (lower panels) parameters for exponential (left panels) and hyperbolic (right panels) collapsing threshold diffusion models.

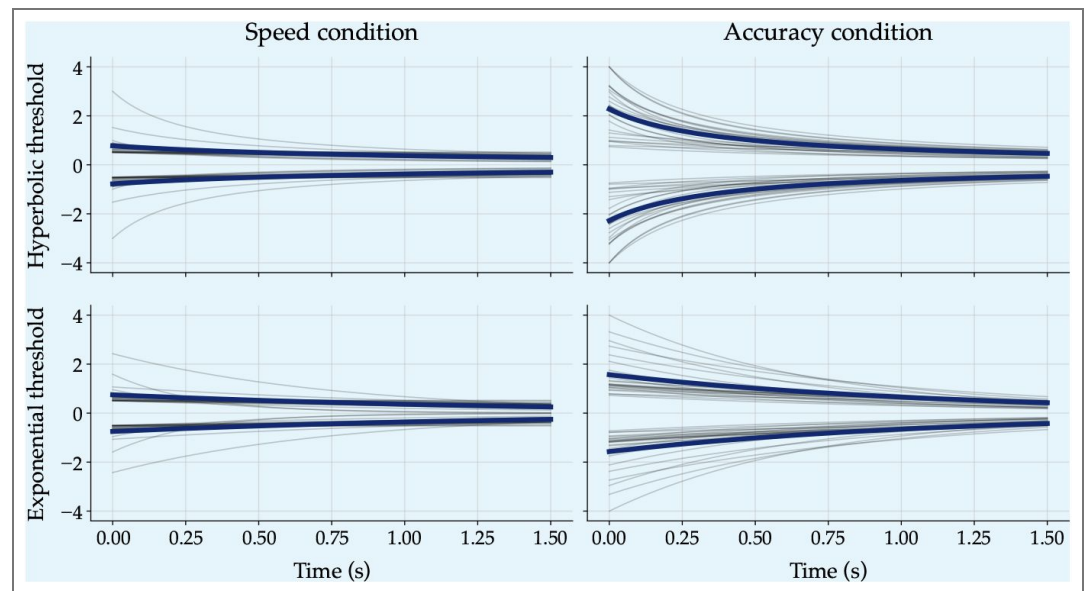
Appendix 6

Threshold estimation for uninformed CT-DDMs

This appendix presents the estimated threshold dynamics for uninformed CT-DDMs. Figure 1 and Figure 2 illustrate the estimated thresholds for each individual in study 1 and study 2, respectively. Similar to the obtained results for NDT-informed CT-DDMs, here we also observed that the decision threshold for most participants collapses largely over time.



Appendix 6—figure 1. The estimated threshold dynamics from uninformed CT-DDMs for individuals in Study 1. The first row illustrates the estimated hyperbolic threshold dynamics, and the second row illustrates the estimated exponential threshold dynamics. The left column shows the threshold dynamics in the speed condition, and the right column shows the accuracy condition. Each gray line corresponds to a subject. The blue line corresponds to the average group level.



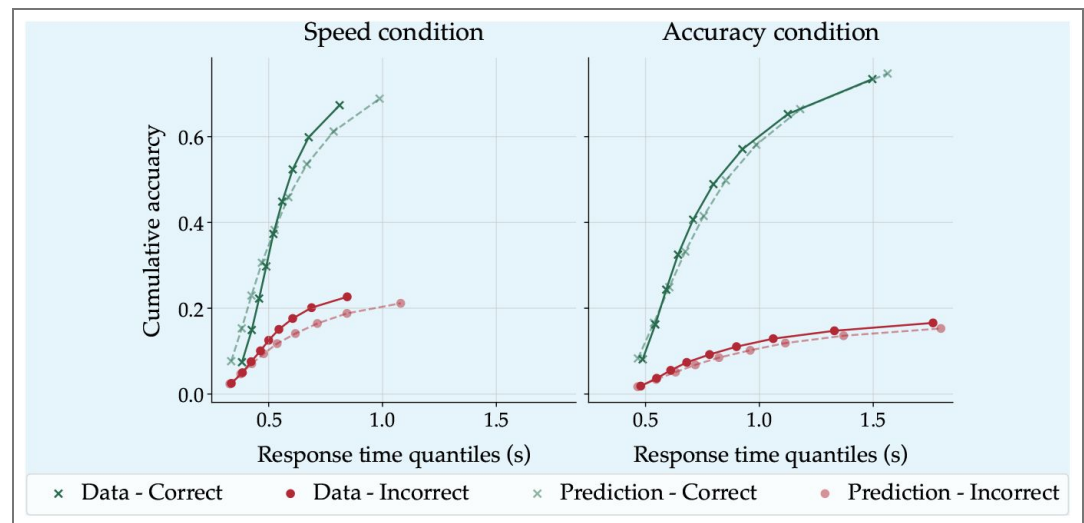
Appendix 6—figure 2. The estimated threshold dynamics from uninformed CT-DDMs for individuals in Study 2. The first row illustrates the estimated hyperbolic threshold dynamics, and the second row illustrates the estimated exponential threshold dynamics. The left column shows the threshold dynamics in the speed condition, and the right column shows the accuracy condition. Each gray line corresponds to a subject. The blue line corresponds to the average group level.

Appendix 7

Results of Fixed-threshold diffusion model with drift variability

Study 1

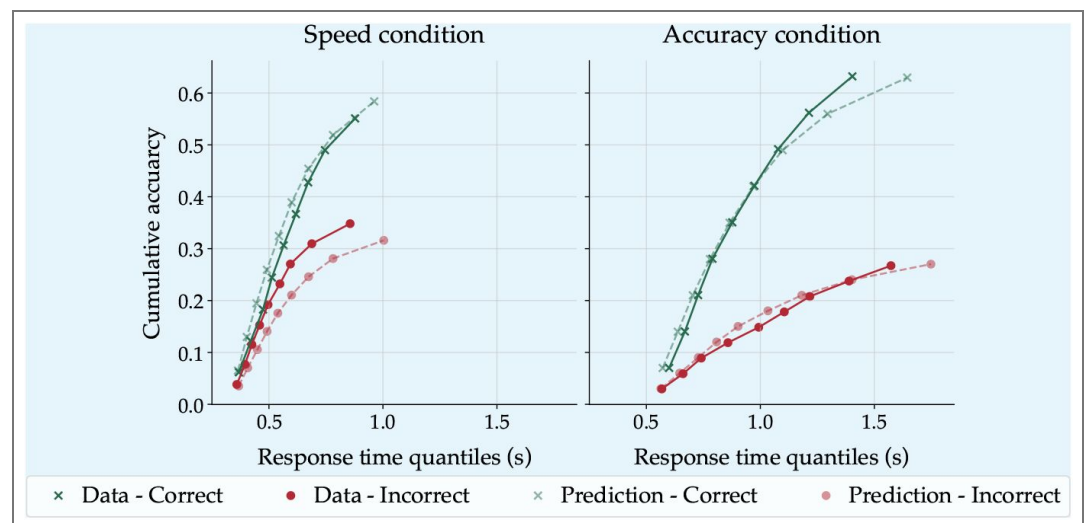
Figure 1 [↗](#) presents the predictions of the uninformed FT-DDM, which includes trial-to-trial variability in drift rate. This model overestimates RT quantiles in the speed condition, particularly for the later response time quantiles for both correct and incorrect responses. However, the predictions of the FT-DDM align more closely with observed empirical data in the accuracy condition. Fitting of this model on the speed condition resulted $BIC_{Behavior} = 5817.66$ and on the accuracy condition $BIC_{Behavior} = 16577.20$. These BIC values are much larger than those of all NDT-informed and uninformed CT-DDMs considered in the main text. This implies that all CT-DDMs perform better than the FT-DDM, which accounts for across-trial variability in drift rate.



Appendix 7—figure 1. Posterior prediction of uninformed FT-DDM against empirical data for speed (left panel) and accuracy (right row) conditions in Study 1. In each panel, the x-axis represents the response time quantiles in seconds, and the y-axis represents the cumulative choice proportion.

Study 2

Similarly, the predictions of the FT-DDM with across-trial variability in drift rate are presented in Figure 2. The FT-DDM overestimates the RT quantiles for both correct and incorrect responses in both speed and accuracy conditions. Moreover, this model underestimates the proportion of incorrect responses chosen in the Speed condition. Also, the model fitting resulted in $BIC_{\text{Behavior}} = 5855.32$, which indicates that the fitting of CT-DDMs considered for study 2 in the main text is better than this model.



Appendix 7—figure 2. Posterior prediction of uninformed FT-DDM against empirical data for speed (left panel) and accuracy (right row) conditions in Study 2. In each panel, the x-axis represents the response time quantiles in seconds, and the y-axis represents the cumulative choice proportion.

Appendix 8

Simulation results for normally distributed non-decision time measurements

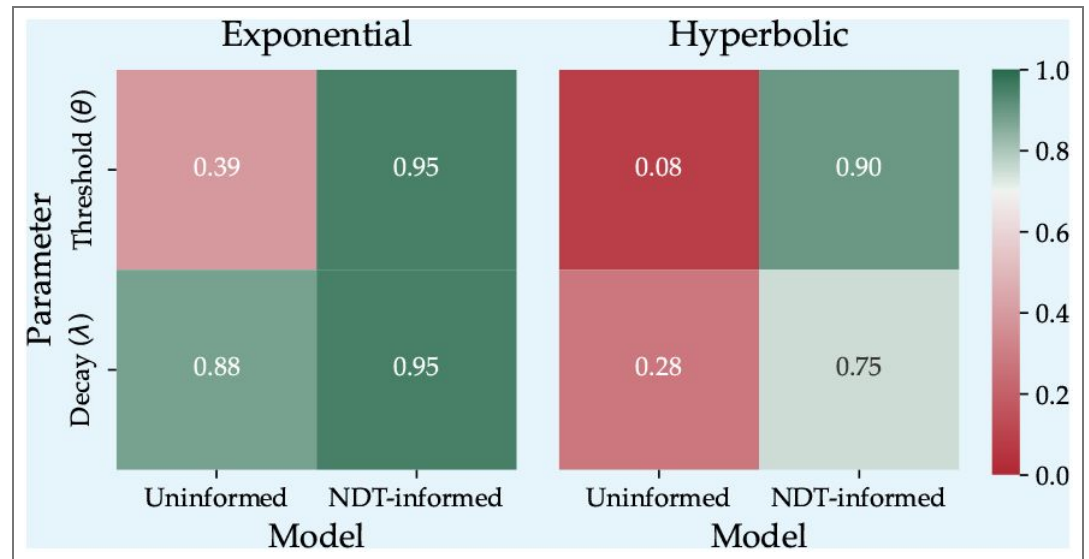
Non-decision time measurements can also be distributed normally around the actual non-decision time. This assumption has been made about the non-decision time variability in the literature (e.g., Ghaderi-Kangavari et al., 2023 [DOI](#); Kira et al., 2025 [DOI](#)). Therefore, it is essential to demonstrate how the main results depend on the distributional assumptions underlying the non-decision time measurements. To this aim, we conducted a simulation study, in which we assumed that the response times and choices are generated from a CT-DDM, and the non-decision time measurements are normally distributed around the mean non-decision time τ with some variance σ_z^2 as follows:

$$\begin{aligned} \text{RT}_n, \text{Choice}_n &\sim \text{CT-DDM}(v, x_0, b_u(t), b_l(t), \tau), \\ Z_n &\sim \mathcal{N}(\tau, \sigma_z^2). \end{aligned} \quad (14)$$

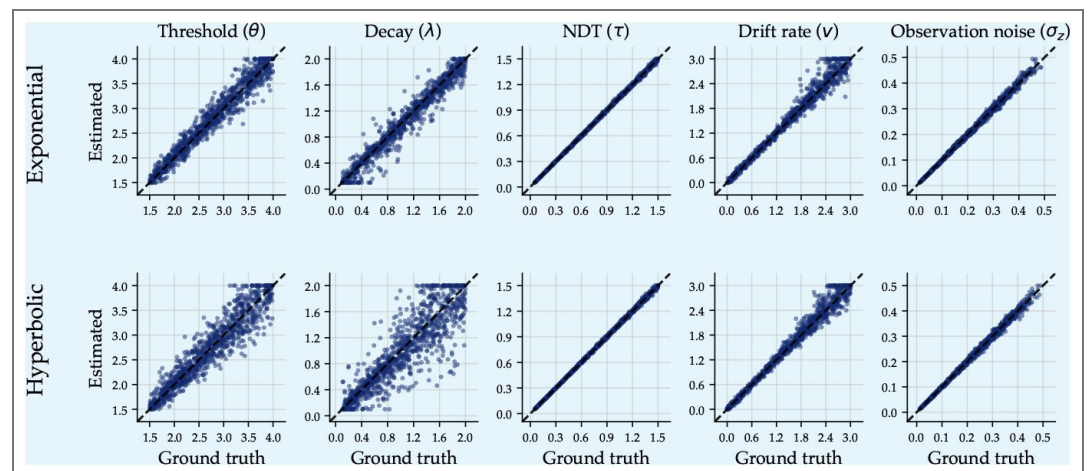
It is important to note that by considering the normal distribution for trial-level non-decision time measurements, the joint log-likelihood of response times, choices, and non-decision measurements becomes as follows:

$$\begin{aligned} -LL_{\text{joint}}(\text{RT}, \text{Choice}, Z \mid v, x_0, b_u(t), b_l(t), \tau, \sigma_z) \\ = -LL_{\text{CT-DDM}}(\text{RT}, \text{Choice} \mid v, x_0, b_u(t), b_l(t), \tau) \\ + \sum_n^N \left[\frac{(Z_n - \tau)^2}{2\sigma_z^2} + \frac{1}{2} \log(2\pi\sigma_z^2) \right]. \end{aligned} \quad (15)$$

Figure 1 [DOI](#) presents the R^2 values for threshold parameters for NDT-informed models with normally distributed non-decision measurements against the uninformed models. Similar to the results reported in the main text with the log-normal assumption for non-decision time measurements, constraining the non-decision time parameter with normally distributed measurements also improves the recovery of the threshold parameters. Also, Figure 3 [DOI](#) presents the estimated parameters against the actual generating parameters for both NDT-informed and uninformed models.

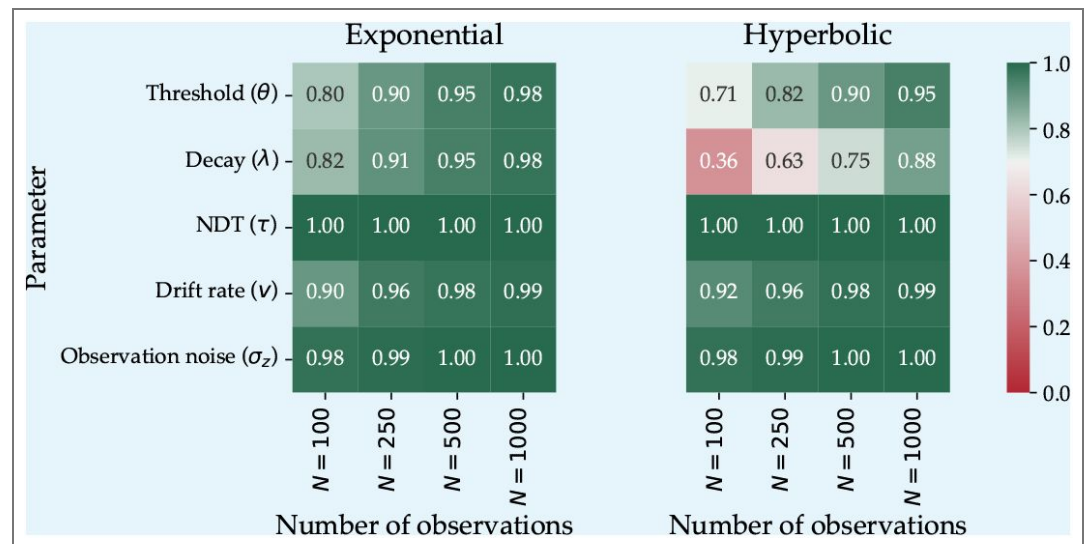


Appendix 8—figure 1. R^2 values measuring the agreement between estimated and ground truth threshold parameters in exponential (left) and hyperbolic (right) collapsing boundary models.



Appendix 8—figure 2. Estimated vs. true parameter values for two modeling approaches. The top row shows parameter recovery for the linear collapsing threshold diffusion model estimated without additional constraints (“Uninformed”), while the bottom row shows recovery for the joint model that incorporates additional non-decision time observations (“NDT informed”). Each point represents a recovered parameter estimate plotted against its corresponding true data-generating value. Dashed diagonal lines indicate perfect recovery.

Figure 3 [illustrates the sensitivity of the parameter estimation to the number of trials](#). Similar to the results for the log-normal distribution, we also observed rapid improvement in parameter recovery with increasing the number of trials.



Appendix 8—figure 3. Illustration of the sensitivity of parameter estimation to the number of trials in the NDT-informed models. R^2 values measure the agreement between estimated and ground truth parameters of the exponential (left) and hyperbolic (right) collapsing threshold models.

Appendix 9

Application to N-dimensional diffusion models

The hyper-spherical diffusion model (HSDM; Smith, 2016; Smith and Corbett, 2019; Hadian Rasanan et al., 2026) is the n -dimensional extension of the well-known DDM (Ratcliff, 1978; Ratcliff and Rouder, 1998). HSDM is an important model to consider because it is applicable in the modeling of decisions with continuous options (e.g., Smith et al., 2020, 2023; Hadian Rasanan et al., 2026), multi-alternative (e.g., Kvam, 2019a; Smith and Corbett, 2019), and multi-attribute decisions (e.g., Smith, 2019; Smith and Corbett, 2019). Thus, it covers a vast range of choice problems (see also Hadian Rasanan et al., 2024a). Recently, Hadian Rasanan et al., 2025 have extended the collapsing threshold version of HSDM and proposed an integral equation method for parameter estimation of this model. The evidence accumulation process in this model can be represented by an n -dimensional vector-valued Wiener process as follows:

$$\begin{bmatrix} dX_1(t) \\ \vdots \\ dX_n(t) \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} dt + \begin{bmatrix} s & & 0 \\ & \ddots & \\ 0 & & s \end{bmatrix} \begin{bmatrix} dW_1(t) \\ \vdots \\ dW_n(t) \end{bmatrix} \quad (16)$$

Hadian Rasanan et al., 2025 showed that the square distance of the an n -dimensional zero-drift process (i.e., $v_i = 0$ for $i = 1, \dots, n$) from the origin (i.e., $Y(t) = X_1^2(t) + \dots + X_n^2(t)$) satisfies the following one-dimensional Feller-type stochastic differential equation (see also Göing-jaesche and Yor, 2003; Kersting et al., 2023):

$$dY(t) = ndt + 2\sqrt{Y(t)}dW(t). \quad (17)$$

In this equation, n (number of dimensions) is the drift rate, and $2\sqrt{Y(t)}$ is the diffusion coefficient. The authors showed that the first-passage time distribution of an n -dimensional zero-drift process starting from y_0 in the presence of a time-dependent threshold $b(t)$ satisfies the following integral equation (Hadian Rasanan et al., 2025):

$$g[b(t), t | y_0, 0] = -2\Psi[b(t), t, | y_0, 0] + 2 \int_0^t g[b(T), T | y_0, 0] \Psi[b(t), t | b(T), T] dT \quad (18)$$

where the kernel function Ψ is defined as follows (Giorno et al., 1989; Hadian Rasanan et al., 2025):

$$\begin{aligned} & \Psi[S(t), t | y, T] \\ &= \frac{1}{2(t-T)} \exp \left\{ -\frac{S(t)+y}{2(t-T)} \right\} \left(\frac{S(t)}{y} \right)^{\frac{n-2}{4}} \left[\left(\frac{d}{dt} S(t) - \frac{S(t)}{t-T} + \mathcal{K}(t) \right) \right. \\ & \quad \left. \times I_{n/2-1} \left(\frac{\sqrt{yS(t)}}{(t-T)} \right) + \frac{\sqrt{yS(t)}}{(t-T)} I_{n/2} \left(\frac{\sqrt{yS(t)}}{(t-T)} \right) \right], \end{aligned} \quad (19)$$

in which $S(t) = b^2(t)$, and $\mathcal{K}(t) = \frac{1}{2}(n-1 - \frac{d}{dt} S(t))$. Also, $I_\zeta(z)$ represents the Bessel function of the second kind of order ζ defined as below (Bell, 2004):

$$I_\zeta(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+\zeta}}{k! \Gamma(k+\zeta+1)}$$

Similar to the one-dimensional CT-DDM, we need to approximate the solution of this integral equation to estimate the first-passage time distribution of the zero-drift process (see Appendix 2). For this purpose, we employed the following discretized scheme, which is identical to the approximation scheme employed for one-dimensional DDM (Buonocore et al., 1987; Hadian Rasanan et al., 2025):

$$\tilde{g}_z[S(\Delta t), \Delta t | y_0, 0] = -2\Psi[S(\Delta t), \Delta t | y_0, 0], \quad (20)$$

$$\begin{aligned} \tilde{g}_z[S(i\Delta t), i\Delta t | y_0, 0] &= -2\Psi[S(i\Delta t), i\Delta t | y_0, 0] \\ &+ 2\Delta t \sum_{j=1}^{i-1} \tilde{g}_z[S(j\Delta t), j\Delta t | y_0, 0] \Psi \\ &[S(i\Delta t), i\Delta t | S(j\Delta t), j\Delta t]. \end{aligned} \quad (21)$$

After estimating the first-passage time distribution of the zero-drift process, we apply the Girsanov change-of-measure theorem (Girsanov, 1960) to obtain the joint distribution of choice and RT for the non-zero drift process (Smith, 2016; Hadian Rasanan et al., 2025). Therefore, the joint distribution of RT and choice can be obtained as follows:

$$\begin{aligned} & P(RT, \varphi_1, \dots, \varphi_{n-1} | b(t), \mathbf{v}, \tau) \\ &= \exp \left[\frac{1}{\sigma^2} (\mathbf{v} \cdot \mathbf{X}(RT - \tau)) - \frac{1}{2\sigma^2} \|\mathbf{v}\|^2 (RT - \tau) \right] \tilde{g}_z(RT - \tau), \end{aligned} \quad (22)$$

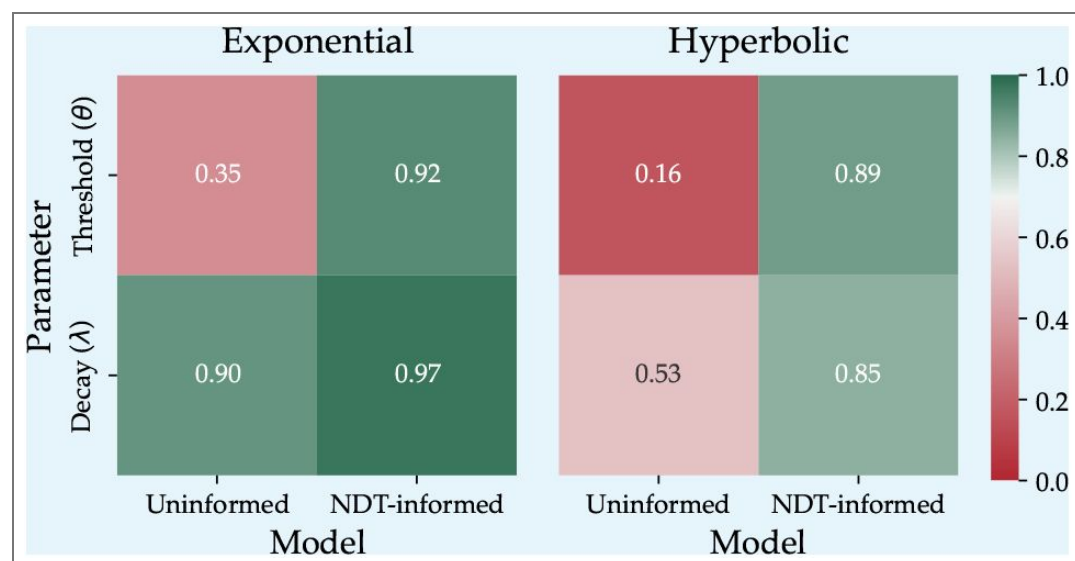
in which $\mathbf{v} = [v_1, \dots, v_n]'$ is the drift vector, ϕ_i ($i = 1, \dots, n-1$) are the response angles with respect to the i -th axis, and $\mathbf{X}(RT - \tau)$ denotes the final position of the accumulator in an n -dimensional space. The following Cartesian representation can be utilized for calculating the elements of the final stopping point $\mathbf{X}(RT - \tau)$:

$$\begin{aligned} X_1(RT - \tau) &= b(RT - \tau) \cos(\varphi_1), \\ X_2(RT - \tau) &= b(RT - \tau) \sin(\varphi_1) \cos(\varphi_2), \\ &\vdots \\ X_{n-1}(RT - \tau) &= b(RT - \tau) \sin(\varphi_1) \sin(\varphi_2) \dots \sin(\varphi_{n-2}) \cos(\varphi_{n-1}), \\ X_n(RT - \tau) &= b(RT - \tau) \sin(\varphi_1) \sin(\varphi_2) \dots \sin(\varphi_{n-2}) \sin(\varphi_{n-1}) \end{aligned}$$

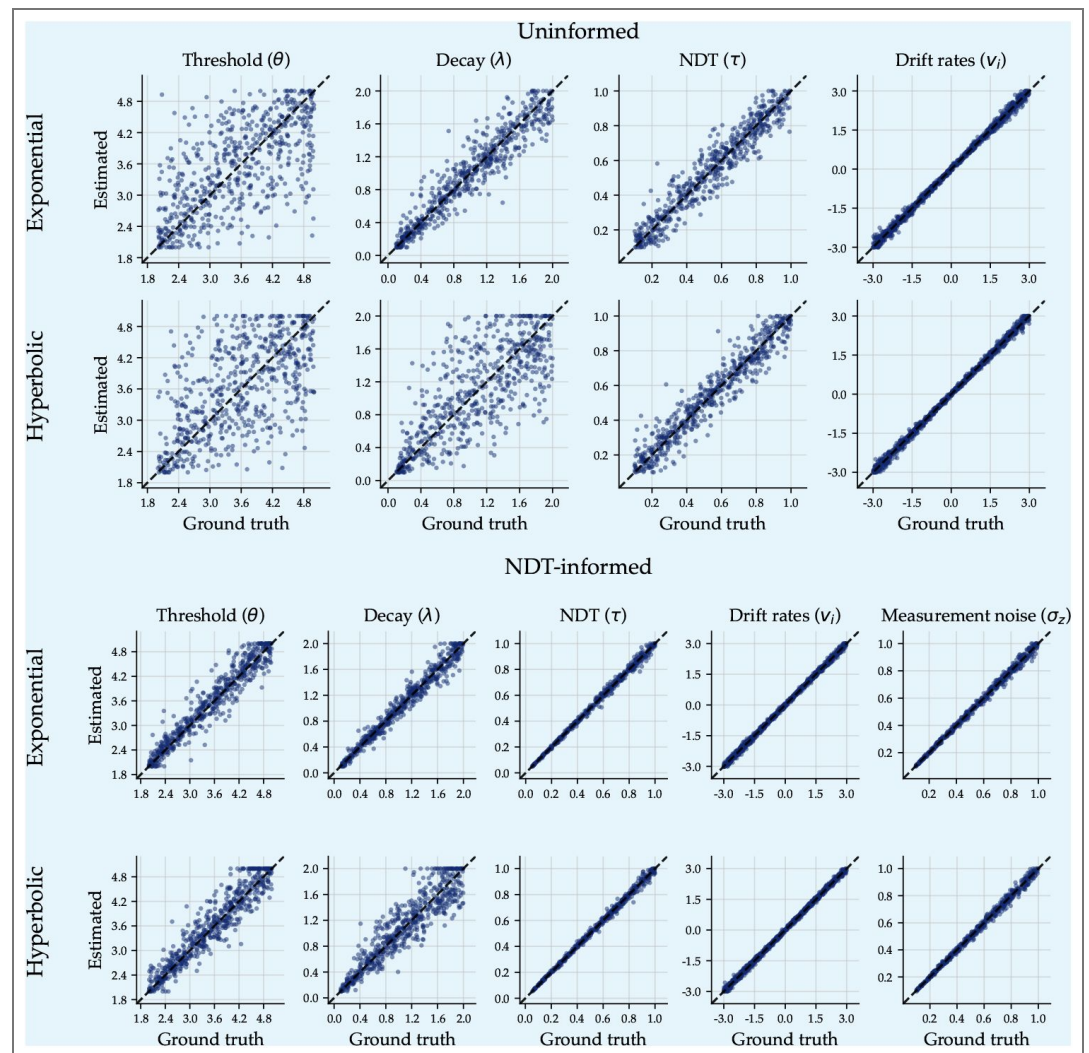
To assess the generalizability of the proposed NDT-informed diffusion modeling, we conducted a parameter recovery study based on the two-dimensional circular diffusion model (Smith, 2016; Hadian Rasanan et al., 2025). Similar to the simulation study reported in the main text, we considered NDT-informed and uninformed collapsing threshold circular diffusion models (CT-CDM) with hyperbolic and exponential collapsing dynamics. We sampled 1000 random parameter sets from the following uniform distributions and then generated 500 trials from CT-CDM for each parameter set:

$$\theta \sim \mathcal{U}[2, 5], \quad \lambda \sim \mathcal{U}[0.1, 2], \quad v_1, v_2 \sim \mathcal{U}[-3, 3], \\ \tau \sim \mathcal{U}[0.05, 1], \quad \sigma \sim \mathcal{U}[0.1, 1].$$

Figures 1 and 2 present the simulation results. Similar to the one-dimensional DDM, we also observe that constraining the non-decision time significantly improves parameter recovery of the two-dimensional CT-CDM.



Appendix 9—figure 1. R^2 values measuring the agreement between estimated and ground truth threshold parameters in exponential (left) and hyperbolic (right) collapsing boundary circular diffusion models.



Appendix 9—figure 2. Estimated vs. true parameter values for two modeling approaches based on the CT-CDM. The top two rows show parameter recovery for models estimated without additional constraints (“Uninformed”), while the bottom two rows show recovery for joint models that incorporate additional non-decision time observations (“NDT-informed”). Each point represents a recovered parameter estimate plotted against its corresponding true data-generating value. Rows correspond to different functional forms (Exponential and Hyperbolic). Dashed diagonal lines indicate perfect recovery.

References

- Anderson JR, Zhang Q, Borst JP, Walsh MM (2016) The discovery of processing stages: Extension of Sternberg’s method. *Psychological Review* **123**:481-509 <https://doi.org/10.1037/rev0000030> | PubMed
- Arnold NR, Bröder A, Bayen UJ (2015) Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research* **79**:882-898 <https://doi.org/10.1007/s00426-014-0608-y> | PubMed
- Bell WW (2004) *Special functions for scientists and engineers* Courier Corporation.
- Berbery HS, van Rijn H, Borst JP (2021) Discovering the brain stages of lexical decision: Behavioral effects originate from a single neural decision process. *Brain and Cognition* **153**:105786 <https://doi.org/10.1016/j.bandc.2021.105786> | PubMed
- Bhui R (2019) Testing optimal timing in value-linked decision making. *Computational Brain & Behavior* **2**:85-94 <https://doi.org/10.1007/s42113-019-0025-9>

- Boehm U, Cox S, Gantner G, Stevenson R (2021) Fast solutions for the first-passage distribution of diffusion models with space-time-dependent drift functions and time-dependent boundaries. *Journal of Mathematical Psychology* **105**:102613 <https://doi.org/10.1016/j.jmp.2021.102613>
- Boehm U, Van Maanen L, Forstmann B, van Rijn H (2014) Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *NeuroImage* **96**:95-105 <https://doi.org/10.1016/j.neuroimage.2014.03.063> | PubMed
- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review* **113**:700-765 <https://doi.org/10.1037/0033-295X.113.4.700> | PubMed
- Bompas A, Sumner P, Hedge C (2025) Non-decision time: The Higgs Boson of decision. *Psychological Review* **132**:330-363 <https://doi.org/10.1037/rev0000487> | PubMed
- Bridwell DA, Cavanagh JF, Collins AG, Nunez MD, Srinivasan R, Stoher S, Calhoun VD (2018) Moving beyond ERP components: a selective review of approaches to integrate EEG and behavior. *Frontiers in Human Neuroscience* **12**:106 <https://doi.org/10.3389/fnhum.2018.00106> | PubMed
- Buonocore A, Giorno V, Nobile AG, Ricciardi LM (1990) On the two-boundary first-crossing-time problem for diffusion processes. *Journal of Applied Probability* **27**:102-114 <https://doi.org/10.2307/3214598>
- Buonocore A, Nobile AG, Ricciardi LM (1987) A new integral equation for the evaluation of first-passage-time probability densities. *Advances in Applied Probability* **19**:784-800 <https://doi.org/10.2307/1427102>
- Burbeck SL, Luce RD (1982) Evidence from auditory simple reaction times for both change and level detectors. *Perception & Psychophysics* **32**:117-133 <https://doi.org/10.3758/BF03204271> | PubMed
- Bürkner PC, Scholz M, Radev ST (2023) Some Models Are Useful, but How Do We Know Which Ones? Towards a Unified Bayesian Model Taxonomy. *Statistics Surveys* **17**:216-310 <https://doi.org/10.1214/23-SS145>
- Busmeyer JR, Gluth S, Rieskamp J, Turner BM (2019) Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in Cognitive Sciences* **23**:251-263 <https://doi.org/10.1016/j.tics.2018.12.003> | PubMed
- Cavanagh JF, Wiecki TV, Kochar A, Frank MJ (2014) Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General* **143**:1476-1488 <https://doi.org/10.1037/a0035813> | PubMed
- Cisek P, Puskas GA, El-Murr S (2009) Decisions in changing conditions: The urgency-gating model. *Journal of Neuro-science* **29**:11560-11571 <https://doi.org/10.1523/JNEUROSCI.1844-09.2009> | PubMed
- Diederich A, Colonius H (2021) A two-stage diffusion modeling approach to the compelled-response task. *Psychological Review* **128**:787-802 <https://doi.org/10.1037/rev0000224> | PubMed
- Diederich A, Oswald P (2016) Multi-stage sequential sampling models with finite or infinite time horizon and variable boundaries. *Journal of Mathematical Psychology* **74**:128-145 <https://doi.org/10.1016/j.jmp.2016.02.010>
- Ditterich J (2006) Evidence for time-variant decision making. *European Journal of Neuroscience* **24**:3628-3641 <https://doi.org/10.1111/j.1460-9568.2006.05221.x> | PubMed
- Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience* **32**:3612-3628 <https://doi.org/10.1523/JNEUROSCI.4010-11.2012> | PubMed
- Dutilh G, Rieskamp J (2016) Comparing perceptual and preferential decision making. *Psychonomic Bulletin & Review* **23**:723-737 <https://doi.org/10.3758/s13423-015-0941-1> | PubMed
- Evans NJ, Bennett AJ, Brown SD (2019) Optimal or not; depends on the task. *Psychonomic Bulletin & Review* **26**:1027-1034 <https://doi.org/10.3758/s13423-018-1536-4> | PubMed

- Evans NJ, Brown SD (2017) People adopt optimal policies in simple decision-making, after practice and guidance. *Psychonomic Bulletin & Review* **24**:597-606 <https://doi.org/10.3758/s13423-016-1135-1> | PubMed
- Evans NJ, Hawkins GE (2019) When humans behave like monkeys: Feedback delays and extensive practice increase the efficiency of speeded decisions. *Cognition* **184**:11-18 <https://doi.org/10.1016/j.cognition.2018.11.014> | PubMed
- Evans NJ, Hawkins GE, Brown SD (2020a) The role of passing time in decision-making. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **46**:316-326 <https://doi.org/10.1037/xlm0000725> | PubMed
- Evans NJ, Trueblood JS, Holmes WR (2020b) A parameter recovery assessment of time-variant models of decisionmaking. *Behavior Research Methods* **52**:193-206 <https://doi.org/10.3758/s13428-019-01218-0> | PubMed
- Fechner GT (1860) *Elemente der psychophysik* **2** Breitkopf u. Härtel.
- Fengler A, Govindarajan LN, Chen T, Frank MJ (2021) Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience. *eLife* **10**:e65074 <https://doi.org/10.7554/eLife.65074> | PubMed
- Fontanesi L, Gluth S, Spektor MS, Rieskamp J (2019) A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin & Review* **26**:1099-1121 <https://doi.org/10.3758/s13423-018-1554-2> | PubMed
- Forstmann BU, Ratcliff R, Wagenmakers EJ (2016) Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology* **67**:641-666 <https://doi.org/10.1146/annurev-psych-122414-033645> | PubMed
- Forstmann BU, Tittgemeyer M, Wagenmakers EJ, Derrfuss J, Imperati D, Brown S (2011) The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience* **31**:17242-17249 <https://doi.org/10.1523/JNEUROSCI.0309-11.2011> | PubMed
- Forstmann BU, Wagenmakers EJ (2015) *An introduction to model-based cognitive neuroscience* **614** Springer.
- Frazier P, Yu AJ (2007) Sequential hypothesis testing under stochastic deadlines. In: NIPS'07: Proceedings of the 21st International Conference on Neural Information Processing Systems. pp. 465-472 <https://dl.acm.org/doi/10.5555/2981562.2981621>
- Fudenberg D, Newey W, Strack P, Strzalecki T (2020) Testing the drift-diffusion model. *Proceedings of the National Academy of Sciences* **117**:33141-33148 <https://doi.org/10.1073/pnas.2011446117> | PubMed
- Fudenberg D, Strack P, Strzalecki T (2018) Speed, accuracy, and the optimal timing of choices. *American Economic Review* **108**:3651-3684 <https://doi.org/10.1257/aer.20150742>
- Ghaderi-Kangavari A, Amani Rad J, Nunez MD (2023) A general integrative neurocognitive modeling framework to jointly describe EEG and decision-making on single trials. *Computational Brain & Behavior* **6**:317-376 <https://doi.org/10.1007/s42113-023-00167-4>
- Ghaderi-Kangavari A, Rad JA, Parand K, Nunez MD (2022) Neuro-cognitive models of single-trial EEG measures describe latent effects of spatial attention during perceptual decision making. *Journal of Mathematical Psychology* **111**:102725 <https://doi.org/10.1016/j.jmp.2022.102725>
- Ging-Jehli NR, Cavanagh JF, Ahn M, Segar DJ, Asaad WF, Frank MJ (2025) Basal ganglia components have distinct computational roles in decision-making dynamics under conflict and uncertainty. *PLoS Biology* **23**:e3002978 <https://doi.org/10.1371/journal.pbio.3002978> | PubMed
- Ging-Jehli NR, Ratcliff R, Arnold LE (2021) Improving neurocognitive testing using computational psychiatry—A systematic review for ADHD. *Psychological Bulletin* **147**:169-231 <https://doi.org/10.1037/bul0000319> | PubMed

- Giorno V**, Nobile AG, Ricciardi LM, Sato S (1989) On the evaluation of first-passage-time probability densities via non-singular integral equations. *Advances in Applied Probability* **21**:20-36 <https://doi.org/10.2307/1427196>
- Girsanov IV** (1960) On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications* **5**:285-301 <https://doi.org/10.1137/1105027>
- Gluth S**, Deakin J, Rieskamp J (2026) A theory of multiattribute search and choice. *Psychological Review* <https://doi.org/10.1037/rev0000614> | PubMed
- Gluth S**, Kern N, Kortmann M, Vitali CL (2020) Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nature Human Behaviour* **4**:634-645 <https://doi.org/10.1038/s41562-020-0822-0> | PubMed
- Gluth S**, Rieskamp J, Büchel C (2012) Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *Journal of Neuroscience* **32**:10686-10698 <https://doi.org/10.1523/JNEUROSCI.0727-12.2012> | PubMed
- Gluth S**, Rieskamp J, Büchel C (2013) Classic EEG motor potentials track the emergence of value-based decisions. *NeuroImage* **79**:394-403 <https://doi.org/10.1016/j.neuroimage.2013.05.005> | PubMed
- Göing-Jaeschke A**, Yor M (2003) A survey and some generalizations of Bessel processes. *Bernoulli* **9**:313-349 <https://doi.org/10.3150/bj/1068128980>
- Gold JI**, Shadlen MN (2007) The neural basis of decision making. *Annual Review of Neuroscience* **30**:535-574 <https://doi.org/10.1146/annurev.neuro.29.051605.113038> | PubMed
- Green DM**, Luce RD (1971) Detection of auditory signals presented at random times: III. *Perception & Psychophysics* **9**:257-268 <https://doi.org/10.3758/BF03212645>
- Grogan J**, Vermeylen L, Mannion S, McCabe C, Monakhovych D, Desender K, O'Connell RG (2025) Neurally-informed modelling unravels a single evidence accumulation process for choices and subsequent confidence reports. *bioRxiv* <https://doi.org/10.1101/2025.06.05.658071>
- Hadian Rasanan AH**, Evans NJ, Amani Rad J, Rieskamp J (2025) Parameter estimation of hyper-spherical diffusion models with a time-dependent threshold: An integral equation method. *Behavior Research Methods* **57**:283 <https://doi.org/10.3758/s13428-025-02810-3> | PubMed
- Hadian Rasanan AH**, Evans NJ, Fontanesi L, Manning C, Huang-Pollock C, Matzke D, Heathcote A, Rieskamp J, Speekenbrink M, Frank MJ, et al. (2024a) Beyond Discrete-Choice Options. *Trends in Cognitive Sciences* **28**:857-870 <https://doi.org/10.1016/j.tics.2024.07.004> | PubMed
- Hadian Rasanan AH**, Evans NJ, Rieskamp J, Amani Rad J (2023) Numerical approximation of the first-passage time distribution of time-varying diffusion decision models: A mesh-free approach. *Engineering Analysis with Boundary Elements* **151**:227-243 <https://doi.org/10.1016/j.enganabound.2023.03.005>
- Hadian Rasanan AH**, Evans NJ, Rieskamp J, Amani Rad J (2024b) Response time and accuracy modeling through the lens of fractional dynamics: A foundational primer. In: Chakraverty S, Jena RM (Eds). *Computation and Modeling for Fractional Order Systems* Elsevier. pp. 1-27 <https://doi.org/10.1016/B978-0-44-315404-1.00006-0>
- Hadian Rasanan AH**, Olschewski S, Rieskamp J (2026) The Projected Spherical Diffusion Model: An Evidence Accumulation Theory for Estimation. *PsyArxiv* https://doi.org/10.31234/osf.io/mhj6v_v1
- Hanks TD**, Mazurek ME, Kiani R, Hopp E, Shadlen MN (2011) Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *Journal of Neuroscience* **31**:6339-6352 <https://doi.org/10.1523/JNEUROSCI.5613-10.2011> | PubMed
- Hawkins GE**, Forstmann BU, Wagenmakers EJ, Ratcliff R, Brown SD (2015) Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience* **35**:2476-2484 <https://doi.org/10.1523/JNEUROSCI.2410-14.2015> | PubMed

- Hawkins GE, Heathcote A (2021) Racing against the clock: Evidence-based versus time-based decisions. *Psychological Review* **128**:222-263 <https://doi.org/10.1037/rev0000259> | PubMed
- Hawkins GE, Mittner M, Forstmann BU, Heathcote A (2019) Modeling distracted performance. *Cognitive Psychology* **112**:48-80 <https://doi.org/10.1016/j.cogpsych.2019.05.002> | PubMed
- Heathcote A, Brown S, Mewhort D (2002) Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review* **9**:394-401 <https://doi.org/10.3758/BF03196299> | PubMed
- Holmes WR (2015) A practical guide to the Probability Density Approximation (PDA) with improved implementation and error characterization. *Journal of Mathematical Psychology* **68**:13-24 <https://doi.org/10.1016/j.jmp.2015.08.006>
- Hübner R, Pelzer T (2020) Improving parameter recovery for conflict drift-diffusion models. *Behavior Research Methods* **52**:1848-1866 <https://doi.org/10.3758/s13428-020-01366-8> | PubMed
- Karalunas SL, Huang-Pollock CL, Nigg JT (2012) Decomposing attention-deficit/hyperactivity disorder (ADHD)-related effects in response speed and variability. *Neuropsychology* **26**:684-694 <https://doi.org/10.1037/a0029936> | PubMed
- Karşılar H, Simen P, Papadakis S, Balci F (2014) Speed accuracy trade-off under response deadlines. *Frontiers in Neuroscience* **8**:248 <https://doi.org/10.3389/fnins.2014.00248> | PubMed
- Katsimpokis D, Hawkins GE, van Maanen L (2020) Not all speed-accuracy trade-off manipulations have the same psychological effect. *Computational Brain & Behavior* **3**:252-268 <https://doi.org/10.1007/s42113-020-00074-y>
- Kelly SP, Corbett EA, O'Connell RG (2021) Neurocomputational mechanisms of prior-informed perceptual decisionmaking in humans. *Nature Human Behaviour* **5**:467-481 <https://doi.org/10.1038/s41562-020-00967-9> | PubMed
- Kersting H, Orvieto A, Proske F, Lucchi A (2023) Mean first exit times of Ornstein–Uhlenbeck processes in high-dimensional spaces. *Journal of Physics A: Mathematical and Theoretical* **56**:215003 <https://doi.org/10.1088/1751-8121/acc559>
- Khodadadi A, Fakhari P, Busemeyer JR (2017) Learning to allocate limited time to decisions with different expected outcomes. *Cognitive Psychology* **95**:17-49 <https://doi.org/10.1016/j.cogpsych.2017.03.002> | PubMed
- Khodadadi A, Townsend JT (2015) On mimicry among sequential sampling models. *Journal of Mathematical Psychology* **68**:37-48 <https://doi.org/10.1016/j.jmp.2015.08.007>
- Kira S, Zylberberg A, Shadlen MN (2025) Incorporation of a cost of deliberation time in perceptual decision making. *Journal of Neuroscience* **45** <https://doi.org/10.1523/JNEUROSCI.2426-24.2025> | PubMed
- Kohfeld DL, Santee JL, Wallace ND (1981) Loudness and reaction time: II Identification of detection components at different intensities and frequencies. *Perception & Psychophysics* **29**:550-562 <https://doi.org/10.3758/BF03207371> | PubMed
- Kohl C, Spieser L, Forster B, Bestmann S, Yarrow K (2020) Centroparietal activity mirrors the decision variable when tracking biased and time-varying sensory evidence. *Cognitive Psychology* **122**:101321 <https://doi.org/10.1016/j.cogpsych.2020.101321> | PubMed
- Krajibich I, Armel C, Rangel A (2010) Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience* **13**:1292-1298 <https://doi.org/10.1038/nn.2635> | PubMed
- Krause J, van Rij J, Borst JP (2024) Word type and frequency effects on lexical decisions are process-dependent and start early. *Journal of Cognitive Neuroscience* **36**:2227-2250 https://doi.org/10.1162/jocn_a_02214 | PubMed
- Kruschke JK, Liddell TM (2018) The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* **25**:178-206 <https://doi.org/10.3758/s13423-016-1221-4> | PubMed

- Kvam PD** (2019a) A geometric framework for modeling dynamic decisions among arbitrarily many alternatives. *Journal of Mathematical Psychology* **91**:14-37 <https://doi.org/10.1016/j.jmp.2019.03.001>
- Kvam PD** (2019b) Modeling accuracy, response time, and bias in continuous orientation judgments. *Journal of Experimental Psychology: Human Perception and Performance* **45**:301-318 <https://doi.org/10.1037/xhp0000606> | PubMed
- Laming DRJ** (1968) *Information theory of choice-reaction times* New York: Wiley.
- Lerche V, Voss A** (2017) Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological Research* **81**:629-652 <https://doi.org/10.1007/s00426-016-0770-5> | PubMed
- Lerche V, Voss A** (2019) Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research* **83**:1194-1209 <https://doi.org/10.1007/s00426-017-0945-8> | PubMed
- Miletić S, Turner BM, Forstmann BU, van Maanen L** (2017) Parameter recovery for the leaky competing accumulator model. *Journal of Mathematical Psychology* **76**:25-50 <https://doi.org/10.1016/j.jmp.2016.12.001>
- Milosavljevic M, Malmaud J, Huth A, Koch C, Rangel A** (2010) The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making* **5**:437-449 <https://doi.org/10.1017/S1930297500001285>
- Moran R** (2015) Optimal decision making in heterogeneous and biased environments. *Psychonomic Bulletin & Review* **22**:38-53 <https://doi.org/10.3758/s13423-014-0669-3> | PubMed
- Murphy PR, Boonstra E, Nieuwenhuis S** (2016) Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nature Communications* **7**:13526 <https://doi.org/10.1038/ncomms13526> | PubMed
- Murrow M, Holmes W** (2024) A parameter recovery assessment of a wide class of evidence accumulation models of decision-making. *Research Square* <https://doi.org/10.21203/rs.3.rs-4722049/v1>
- Murrow M, Holmes WR** (2024) PyBEAM: A Bayesian approach to parameter inference for a wide class of binary evidence accumulation models. *Behavior Research Methods* **56**:2636-2656 <https://doi.org/10.3758/s13428-023-02162-w> | PubMed
- Nejati V, Hadian Rasanah AH, Amani Rad J, Alavi MM, Haghi S, Nitsche MA** (2022) Transcranial direct current stimulation (tDCS) alters the pattern of information processing in children with ADHD: Evidence from drift diffusion modeling. *Neurophysiologie Clinique* **52**:17-27 <https://doi.org/10.1016/j.neucli.2021.11.005> | PubMed
- Nunez MD, Gosai A, Vandekerckhove J, Srinivasan R** (2019) The latency of a visual evoked potential tracks the onset of decision making. *NeuroImage* **197**:93-108 <https://doi.org/10.1016/j.neuroimage.2019.04.052> | PubMed
- Nunez MD, Schubert AL, Frischkorn GT, Oberauer K** (2025) Cognitive models of decision-making with identifiable parameters: Diffusion Decision Models with within-trial noise. *Journal of Mathematical Psychology* **125**:102917 <https://doi.org/10.1016/j.jmp.2025.102917>
- Nunez MD, Vandekerckhove J, Srinivasan R** (2017) How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology* **76**:117-130 <https://doi.org/10.1016/j.jmp.2016.03.003> | PubMed
- O'Connell RG, Dockree PM, Kelly SP** (2012) A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience* **15**:1729-1735 <https://doi.org/10.1038/nn.3248> | PubMed
- Olschewski S, Mullett TL, Stewart N** (2025) Optimal allocation of time in risky choices under opportunity costs. *Cognitive Psychology* **157**:101716 <https://doi.org/10.1016/j.cogpsych.2025.101716> | PubMed

- Palestro JJ, Weichart E, Sederberg PB, Turner BM (2018) Some task demands induce collapsing bounds: Evidence from a behavioral analysis. *Psychonomic Bulletin & Review* **25**:1225-1248 <https://doi.org/10.3758/s13423-018-1479-9> | PubMed
- Pedersen ML, Frank MJ, Biele G (2017) The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review* **24**:1234-1251 <https://doi.org/10.3758/s13423-016-1199-y> | PubMed
- Pirrone A, Dickinson A, Gomez R, Stafford T, Milne E (2017) Understanding perceptual judgment in autism spectrum disorder using the drift diffusion model. *Neuropsychology* **31**:173-180 <https://doi.org/10.1037/neu0000320> | PubMed
- Radev ST, Mertens UK, Voss A, Ardizzone L, Köthe U (2020) BayesFlow: Learning complex stochastic models with invertible neural networks. *IEEE Transactions on Neural Networks and Learning Systems* **33**:1452-1466 <https://doi.org/10.1109/TNNLS.2020.3042395> | PubMed
- Radev ST, Schmitt M, Pratz V, Picchini U, Köthe U, Bürkner PC (2023a) JANA: Jointly amortized neural approximation of complex Bayesian models. In: Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence. pp. 1695-1706 <https://proceedings.mlr.press/v216/radev23a.html>
- Radev ST, Schmitt M, Schumacher L, Elsemüller L, Pratz V, Schälte Y, Köthe U, Bürkner PC (2023b) BayesFlow: Amortized Bayesian Workflows With Neural Networks. *Journal of Open Source Software* **8**:5702 <https://doi.org/10.21105/joss.05702>
- Ratcliff R (1978) A theory of memory retrieval. *Psychological Review* **85**:59-108 <https://doi.org/10.1037/0033-295X.85.2.59>
- Ratcliff R (2008) Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging* **23**:900-916 <https://doi.org/10.1037/a0013930> | PubMed
- Ratcliff R (2022) Integrated diffusion models for distance effects in number memory. *Cognitive Psychology* **138**:101516 <https://doi.org/10.1016/j.cogpsych.2022.101516> | PubMed
- Ratcliff R, Frank MJ (2012) Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. *Neural Computation* **24**:1186-1229 https://doi.org/10.1162/NECO_a_00270 | PubMed
- Ratcliff R, McKoon G (2008) The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation* **20**:873-922 <https://doi.org/10.1162/neco.2008.12-06-420> | PubMed
- Ratcliff R, McKoon G (2018) Modeling numerosity representation with an integrated diffusion model. *Psychological Review* **125**:183-217 <https://doi.org/10.1037/rev0000085> | PubMed
- Ratcliff R, McKoon G (2023) Reexamining the effects of speed-accuracy instructions with a diffusion-model-based analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **49**:1732-1751 <https://doi.org/10.1037/xlm0001285> | PubMed
- Ratcliff R, Rouder JN (1998) Modeling response times for two-choice decisions. *Psychological Science* **9**:347-356 <https://doi.org/10.1111/1467-9280.00067>
- Ratcliff R, Smith PL (2004) A comparison of sequential sampling models for two-choice reaction time. *Psychological Review* **111**:333-367 <https://doi.org/10.1037/0033-295X.111.2.333> | PubMed
- Ratcliff R, Smith PL, Brown SD, McKoon G (2016) Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences* **20**:260-281 <https://doi.org/10.1016/j.tics.2016.01.007> | PubMed
- Ratcliff R, Thapar A, McKoon G (2010) Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology* **60**:127-157 <https://doi.org/10.1016/j.cogpsych.2009.09.001> | PubMed
- Ratcliff R, Tuerlinckx F (2002) Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review* **9**:438-481 <https://doi.org/10.3758/BF03196302> | PubMed
- Richter T, Ulrich R, Janczyk M (2023) Diffusion models with time-dependent parameters: An analysis of computational effort and accuracy of different numerical methods. *Journal of Mathematical Psychology* **114**:102756 <https://doi.org/10.1016/j.jmp.2023.102756>

- Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience* **22**:9475-9489 <https://doi.org/10.1523/JNEUROSCI.22-21-09475.2002> | PubMed
- Schall JD (2004) On building a bridge between brain and behavior. *Annual Review of Psychology* **55**:23-50 <https://doi.org/10.1146/annurev.psych.55.090902.141907> | PubMed
- Schmiedek F, Oberauer K, Wilhelm O, Süß HM, Wittmann WW (2007) Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General* **136**:414-429 <https://doi.org/10.1037/0096-3445.136.3.414> | PubMed
- Schubert AL, Frischkorn GT (2020) Neurocognitive psychometrics of intelligence: How measurement advancements unveiled the role of mental speed in intelligence differences. *Current Directions in Psychological Science* **29**:140-146 <https://doi.org/10.1177/0963721419896365>
- Servant M, Logan GD, Gajdos T, Evans NJ (2021) An integrated theory of deciding and acting. *Journal of Experimental Psychology: General* **150**:2435-2454 <https://doi.org/10.1037/xge0001063> | PubMed
- Servant M, White C, Montagnini A, Burle B (2016) Linking theoretical decision-making mechanisms in the Simon task with electrophysiological data: A model-based neuroscience study in humans. *Journal of Cognitive Neuroscience* **28**:1501-1521 https://doi.org/10.1162/jocn_a_00989 | PubMed
- Sheu CF, Ratcliff R (1995) The application of Fourier deconvolution to reaction time data: A cautionary note. *Psychological Bulletin* **118**:285-299 <https://doi.org/10.1037/0033-2909.118.2.285> | PubMed
- Shinn M, Lam NH, Murray JD (2020) A flexible framework for simulating and fitting generalized drift-diffusion models. *eLife* **9**:e56938 <https://doi.org/10.7554/eLife.56938> | PubMed
- Smith PL (1990) Obtaining meaningful results from Fourier deconvolution of reaction time data. *Psychological Bulletin* **108**:533-550 <https://doi.org/10.1037/0033-2909.108.3.533> | PubMed
- Smith PL (2000) Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology* **44**:408-463 <https://doi.org/10.1006/jmps.1999.1260> | PubMed
- Smith PL (2016) Diffusion theory of decision making in continuous report. *Psychological Review* **123**:425-451 <https://doi.org/10.1037/rev0000023> | PubMed
- Smith PL (2019) Linking the diffusion model and general recognition theory: Circular diffusion with bivariate normally distributed drift rates. *Journal of Mathematical Psychology* **91**:145-158 <https://doi.org/10.1016/j.jmp.2019.06.002>
- Smith PL (2023) “Reliable organisms from unreliable components” revisited: The linear drift, linear infinitesimal variance model of decision making. *Psychonomic Bulletin & Review* **30**:1323-1359 <https://doi.org/10.3758/s13423-022-02237-3> | PubMed
- Smith PL, Corbett EA (2019) Speeded multielement decision-making as diffusion in a hypersphere: Theory and application to double-target detection. *Psychonomic Bulletin & Review* **26**:127-162 <https://doi.org/10.3758/s13423-018-1491-0> | PubMed
- Smith PL, Corbett EA, Lilburn SD (2023) Diffusion theory of the antipodal “shadow” mode in continuous-outcome, coherent-motion decisions. *Psychological Review* **130**:1167-1202 <https://doi.org/10.1037/rev0000377> | PubMed
- Smith PL, Ratcliff R (2022) Modeling evidence accumulation decision processes using integral equations: Urgencygating and collapsing boundaries. *Psychological Review* **129**:235-267 <https://doi.org/10.1037/rev0000301> | PubMed
- Smith PL, Saber S, Corbett EA, Lilburn SD (2020) Modeling continuous outcome color decisions with the circular diffusion model: Metric and categorical properties. *Psychological Review* **127**:562-590 <https://doi.org/10.1037/rev0000185> | PubMed
- Starns JJ, Ratcliff R (2010) The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging* **25**:377-390 <https://doi.org/10.1037/a0018022> | PubMed

- Steinemann NA, O'Connell RG, Kelly SP (2018) Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications* **9**:3627 <https://doi.org/10.1038/s41467-018-06117-0> | PubMed
- Stone M (1960) Models for choice-reaction time. *Psychometrika* **25**:251-260 <https://doi.org/10.1007/BF02289729>
- Tajima S, Drugowitsch J, Patel N, Pouget A (2019) Optimal policy for multi-alternative decisions. *Nature Neuroscience* **22**:1503-1511 <https://doi.org/10.1038/s41593-019-0453-9> | PubMed
- Tajima S, Drugowitsch J, Pouget A (2016) Optimal policy for value-based decision-making. *Nature Communications* **7**:12400 <https://doi.org/10.1038/ncomms12400> | PubMed
- Thura D, Beauregard-Racine J, Fradet CW, Cisek P (2012) Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology* **108**:2912-2930 <https://doi.org/10.1152/jn.01071.2011> | PubMed
- Trueblood JS, Heathcote A, Evans NJ, Holmes WR (2021) Urgency, leakage, and the relative nature of information processing in decision-making. *Psychological Review* **128**:160-186 <https://doi.org/10.1037/rev0000255> | PubMed
- Turner BM, Forstmann BU, Love BC, Palmeri TJ, Van Maanen L (2017) Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology* **76**:65-79 <https://doi.org/10.1016/j.jmp.2016.01.001> | PubMed
- Turner BM, Sederberg PB (2014) A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review* **21**:227-250 <https://doi.org/10.3758/s13423-013-0530-0> | PubMed
- Turner BM, Van Maanen L, Forstmann BU (2015) Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. *Psychological Review* **122**:312-336 <https://doi.org/10.1037/a0038894> | PubMed
- Van Maanen L, Portoles O, Borst JP (2021) The discovery and interpretation of evidence accumulation stages. *Computational Brain & Behavior* **4**:395-415 <https://doi.org/10.1007/s42113-021-00105-2>
- Verdonck S, Tuerlinckx F (2016) Factoring out nondecision time in choice reaction time data: Theory and implications. *Psychological Review* **123**:208-2018 <https://doi.org/10.1037/rev0000019> | PubMed
- von Krause M, Radev ST, Voss A (2022) Mental Speed Is High until Age 60 as Revealed by Analysis of over a Million Participants. *Nature Human Behaviour* **6**:700-708 <https://doi.org/10.1038/s41562-021-01282-7> | PubMed
- Voskuilen C, Ratcliff R, Smith PL (2016) Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology* **73**:59-79 <https://doi.org/10.1016/j.jmp.2016.04.008> | PubMed
- Voss A, Rothermund K, Voss J (2004) Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition* **32**:1206-1220 <https://doi.org/10.3758/BF03196893> | PubMed
- Voss A, Voss J (2008) A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology* **52**:1-9 <https://doi.org/10.1016/j.jmp.2007.09.005>
- Wazwaz AM (2011) *Linear and nonlinear integral equations* **639** Springer.
- Weindel G (2021) On the measurement and estimation of cognitive processes with electrophysiological recordings and reaction time modeling. Aix-Marseille.
- Weindel G, Anders R, Alario F, Burle B (2021a) Assessing model-based inferences in decision making with single-trial response time decomposition. *Journal of Experimental Psychology: General* **150**:1528-1555 <https://doi.org/10.1037/xge0001010> | PubMed
- Weindel G, Borst JP, van Maanen L (2025) Decision-making components and times revealed by the single-trial electroencephalogram. *eLife* **14**:RP108049 <https://doi.org/10.7554/eLife.108049.3>
- Weindel G, Gajdos T, Burle B, Alario FX (2021b) The decisive role of non-decision time for interpreting the parameters of decision making models. *PsyArXiv* <https://doi.org/10.31234/osf.io/gewb3>

- Weindel G, van Maanen L, Borst JP (2024) Trial-by-trial detection of cognitive events in neural time-series. *Imaging Neuroscience* **2**:1-28 https://doi.org/10.1162/imag_a_00400 | PubMed
- White CN, Servant M, Logan GD (2018) Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic Bulletin & Review* **25**:286-301 <https://doi.org/10.3758/s13423-017-1271-2> | PubMed
- Yap MJ, Balota DA, Sibley DE, Ratcliff R (2012) Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance* **38**:53-79 <https://doi.org/10.1037/a0024177> | PubMed
- Zhang S, Lee MD, Vandekerckhove J, Maris G, Wagenmakers EJ (2014) Time-varying boundaries for diffusion models of decision making and response time. *Frontiers in Psychology* **5**:1364 <https://doi.org/10.3389/fpsyg.2014.01364> | PubMed
- Zhao WJ, Walasek L, Bhatia S (2020) Psychological mechanisms of loss aversion: A drift-diffusion decomposition. *Cognitive Psychology* **123**:101331 <https://doi.org/10.1016/j.cogpsych.2020.101331> | PubMed
- Weindel, Gabriel and Borst, Jelmer P. and van Maanen, Leendert (2025) Decision-making components and times revealed by the single-trial electro-encephalogram. GitHub. ID GWeindel/decision-times <https://github.com/GWeindel/decision-times>
- Boehm, Udo and Van Maanen, Leendert and Forstmann, Birte and van Rijn, Hedderik (2014) Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. GitHub. ID GWeindel/man_hmp https://github.com/GWeindel/man_hmp

Peer reviews

Reviewer #1 (Public review):

Summary:

This paper proposes a non-decision time (NDT)-informed approach to estimating time-varying decision thresholds in diffusion models of decision making. The manuscript motivates the method well, outlines the identifiability issues it is intended to address, and evaluates it using simulations and two empirical datasets. The aim is clear, the scope is deliberately focused, and the manuscript is well written. The core idea is interesting, technically grounded, and a meaningful contribution to ongoing work on collapsing thresholds.

Strengths:

The manuscript is logically structured and easy to follow. The emphasis on parameter recovery is appropriate and appreciated. The finding that the exponential NDT-informed function produces substantially better recovery than the hyperbolic form is useful, given the importance placed on identifiability earlier in the paper. The threshold visualisations are also helpful for interpreting what the models are doing. Overall, the work offers a well-defined, methodologically oriented contribution that will interest researchers working on time-varying thresholds.

Weaknesses / Areas for Clarification:

A few points would benefit from additional clarification following the previous revision:

Returning to one point from my original review. The applications to empirical data describe 6 models, including the FT-DDM with across-trial variability described as a benchmark. Yet the modelling results for both studies (Tables 2 and 3) show only 5 models, omitting the benchmark model. I acknowledge the comment about this in the response letter (no other FT-DDM models in the main text have across-trial variability). Nevertheless, for this to serve as a

benchmark, the modelling results really should be reported in Tables 2 and 3 with the other 5 models, and the model goodness of fit figures currently shown in Appendix 7 incorporated into Figures 10 and 13 of the main text. Unfortunately, as it currently reads, it looks as if something is being obscured, which I don't believe is the intention. This won't change the primary conclusions, but it will increase transparency and ease of interpretation.

Many thanks for introducing Appendix 1 to the revised manuscript. Additional motivation for the central issue is always helpful. However, I'm not (yet) convinced the simulation study in Appendix 1 achieves the intended aim. The simulation study shows that NDT misspecification leads to poorer recovery of other CT parameters (i.e., if a generating value is perturbed and then held fixed at that perturbed value during estimation, recovery of the remaining parameters deteriorates). This demonstrates the consequences of fixing NDT incorrectly, but it does not seem to address the central claim of the manuscript: that NDT and the CT threshold parameters trade off when they are estimated simultaneously. I had expected the simulation study to estimate NDT alongside the remaining model parameters, mirroring the estimation procedure used in the empirical analyses. Such a simulation would directly test whether the two sets of parameters compensate for one another during estimation, and whether this results in poor recovery of the NDT parameters.

<https://doi.org/10.7554/eLife.109730.2.sa3>

Reviewer #2 (Public review):

Summary:

The authors use simulations and empirical data fitting in order to demonstrate that informing a decision model using noisy single-trial estimates of an underlying fixed non-decision time can guide the model to more reliable parameter estimates, especially when the model has collapsing bounds.

Strengths:

The paper is well written and motivated, with clear depth of knowledge in the areas of neurophysiology of decision-making, sequential sampling models, and in particular, the phenomenon of collapsing decision bounds.

Two large-scale simulations are run to test parameter recovery, and two empirical datasets are fit and assessed; the fitting procedures themselves are state-of-the-art, and the study makes use of a very new and well-designed ERP decomposition algorithm that provides single-trial estimates of the duration of diffusion; the results provide inferences about the operation of decision bound collapse - all of this is impressive.

Weaknesses:

This is an interesting and promising idea, but a very important issue is not clear: it is an intuitive principle that information from an external empirical source can enhance the reliability of parameter estimates for a given model, but how can the overall BIC improve, unless it is in fact a different model?

Comment on revised version.

Thanks to the authors for their responses and inclusion of additional analyses and simulations. Thanks, in particular for clarifying a crucial detail, that the ndt-informed model actually assumes, like the uninformed model, that there is no variability in the non-decision time, and the idea is that the variable single-trial measurements of non-decision time are noisy estimates of an underlying, constant ndt. The revised paper itself has not made this clear - for example, throughout the Intro, there is no statement that the behavioural model

assumes a trial-invariant ndt , and line 231 still calls τ the 'mean' non decision time, implying there is a distribution rather than an invariant single value in the behavioural model.

One implication of the above is that if the lognormal σ is purely measurement noise that does not relate to actual variation in the underlying decision process generating behaviour, then the HMP latencies should not relate to behaviour, e.g. shorter latencies predicting shorter RT. I assume that even if the authors did find such a relationship, the principle still stands that a model with fixed ndt is more accurately fit when there are single-trial ndt estimates whose mean provides a constraint on that ndt value, than without such measurements. Still, given ndt variability is a core feature of many decision models, the authors could comment on whether the strategy would work in theory for a model with ndt variability (in the behavioural part), where the single trial estimates would then presumably reflect a mix of measurement noise and genuine ndt variability.

Another more important implication is that since it is in fact the same model being compared with and without the HMP data guiding the fixed ndt estimate, the reason the fit quality improves with HMP-information is not because it is a better model per se (it is the same model) but because without the HMP guidance, the search algorithm somehow gets lost and fails to find the 'optimal' parameter vector. That is, the parameter vector (just the parameters that relate to the behavioural model itself, not the HMP lognormally-distributed noise associated with VEP measurements) identified as optimal in the HMP-informed version of the model exists in the parameter space of the model without HMP information, but it is just not found? I raised this implication before, and it is still not clear whether it applies. I'm sorry to press on it, but it is critical for readers to understand why it is that neural information can improve overall model fit. Again, the enhancement of parameter recovery (like in Nunez 2025) makes sense, but the enhancement of the "model's fit to behavioural data" does not, without pointing to a deficiency in the search algorithm / fitting procedure.

The authors state in their replies that the onset of bound collapse is set at accumulation onset and imply that setting it instead at stimulus onset could "mathematically resolve the issue" but they don't do it because it is implausible. It is in fact not only plausible but clearly evidenced in empirical data - collapsing bounds are implemented neurally through urgency signals, and these can begin to dynamically build toward threshold well before, let alone at, stimulus onset. There is nothing bizarre about this - we can prepare movements without sensory input, and indeed even if choosing actions based on a sensory discrimination, motor preparation can launch well before the sensory evidence (e.g. Stanford, Salinas et al 2010) and this in effect collapses the bound on cumulative evidence for triggering action before any evidence actually arrives. So, Urgency/bound-collapse does not need to be triggered by a stimulus; it can start in anticipation of the stimulus. It seems critical, therefore, for the authors to clarify this point - does re-defining the onset of the collapse at stimulus onset remove the trade-off and render unnecessary the neurally-informed ndt estimation?

Related to this, it is still not clear how bias in the estimation of nondecision time would not be a problem. What if, for example, it is the end of the N2 rather than the peak of the N2 that marks accumulation onset, and/or there is an additional fixed motor time that adds to the N2-based marker to make the full nondecision time that applies in the underlying decision process. By definition (and I think this is essentially what the authors' new simulations verify), because of the trade-offs, this bias would simply be absorbed in shifted estimates of θ and λ describing the bound collapse function. But wasn't the whole point of the exercise to more accurately estimate those parameters? The obvious implication is that the parameter-estimation accuracy of the ERP-informed model is determined by the accuracy with which the proposed ERP marker directly pinpoints the full nondecision time without bias, but this is not at all obvious in the paper as written. Importantly, in the example scenario I describe above where accumulation onsets when N2 ends, there may still be a

perfect correlation of N2 peak latency with underlying ndt across trials - they could still be very strongly "linked" statistically, but we can't know what size offset might be involved.

<https://doi.org/10.7554/eLife.109730.2.sa2>

Reviewer #3 (Public review):

The current paper addresses an important issue in evidence accumulation models: many modelers implement flat decision boundaries because the collapsing alternatives are hard to reliably estimate. Here, using simulations the authors demonstrate that parameter recovery can be drastically improved by providing the model with additional data (specifically, an EEG-informed estimate of non-decision time). Moreover, in two empirical datasets it is shown that those EEG-informed models provide a better fit to the data. The method seems sound and promising and might inform future work on the debate regarding flat vs collapsing choice boundaries. As an evidence-accumulation enthusiast, I am quite excited about this work, although for the more broader audience the immediate applicability of this approach seems limited because it does require EEG data (i.e. limiting widespread use of the method or e.g. answering questions about individual differences that require very large N).

Comments on revised version.

I thought the authors carefully addressed my concerns.

<https://doi.org/10.7554/eLife.109730.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Reviewer #1 (Public review):

Summary:

This paper proposes a non-decision time (NDT)-informed approach to estimating timevarying decision thresholds in diffusion models of decision making. The manuscript motivates the method well, outlines the identifiability issues it is intended to address, and evaluates it using simulations and two empirical datasets. The aim is clear, the scope is deliberately focused, and the manuscript is well written. The core idea is interesting, technically grounded, and a meaningful contribution to ongoing work on collapsing thresholds.

Strengths:

The manuscript is logically structured and easy to follow. The emphasis on parameter recovery is appropriate and appreciated. The finding that the exponential NDT-informed function produces substantially better recovery than the hyperbolic form is useful, given the importance placed on identifiability earlier in the paper. The threshold visualisations are also helpful for interpreting what the models are doing. Overall, the work offers a well-defined, methodologically oriented contribution that will interest researchers working on time-varying thresholds.

We appreciate the positive and constructive feedback. We have addressed your comments in the revised manuscript, as detailed in our responses to the specific comments below.

Weaknesses / Areas for Clarification:

A few points would benefit from clarification, additional analysis, or revised presentation:

(1) It would help readers to see a concrete demonstration of the trade-off between NDT and collapsing thresholds, to give a sense of the scale of the identifiability problem motivating the work.

Thank you for this constructive suggestion. We conducted a new simulation study in which we considered the non-decision time as a fixed parameter and estimated the remaining parameters of the collapsing threshold model. In this simulation study, we contaminated the non-decision time with noise at six levels (i.e., 0%, 2%, 4%, 6%, 8%, and 10%). The simulation results showed that increasing the noise level in non-decision time worsens the estimation of the starting threshold and decay rate parameters, indicating a trade-off between non-decision time and collapsing threshold parameters. The results of this simulation are presented in Appendix 1 in the new version of the manuscript.

(2) Before moving to the empirical datasets, the manuscript really needs a simulation-based model recovery comparison, since all major conclusions of the empirical applications rely on model comparison. One approach might be to simulate from (a) an FT model with across-trial drift variability and (b) one of the CT models, then fit both models to each of the simulated data sets. This would address a longstanding issue: sometimes CT models are preferred even when the estimated collapse in the thresholds is close to zero. A recovery study would confirm that model selection behaves sensibly in the new framework.

We are grateful for this constructive comment. The revised manuscript includes a model recovery study (see Simulation study 2, in particular Table 1 and the accompanying text). In this simulation, as the reviewer suggested, we generated data from FT-DDM with across-trial variability in drift rate and CT-DDM with hyperbolic and exponential collapsing thresholds. Then, for each generated dataset, we fitted two models and classified the datasets based on goodness-of-fit and estimated decay rate. The results show that incorporating the decay rate, which can be reliably estimated by the NDT-informed modeling framework, significantly improves the precision of model recovery.

(3) An additional subtle point is that BIC is defined in terms of the maximised log-likelihood of the model for the data being modelled. In the joint model, the parameter estimates maximise the combined likelihood of behavioural and non-decision-time data. This means the behavioural log-likelihood evaluated at the joint MLEs is not the behavioural MLE. If BIC is being computed for the behavioural data only, this breaks the assumptions underlying BIC. The only valid BIC here would be one defined for the joint model using the joint likelihood.

We thank the reviewer for raising this important methodological point. We agree that, strictly speaking, the behavioral log-likelihood evaluated at the joint maximum likelihood estimates is not guaranteed to be equal to the behavioral maximum likelihood estimate. Therefore, if one were to interpret BIC_{Behavior} for joint (NDT-informed) models as a conventional BIC derived from a purely behavioral maximum likelihood fit, this would indeed violate the standard assumptions underlying BIC. Our intention, however, was not to claim that BIC_{Behavior} represents a formally valid BIC in the strict information-theoretic sense. Rather, we used it as a diagnostic measure to assess how well the jointly estimated parameters account for the behavioral data relative to the uninformed models. Importantly, in our datasets, the behavioral log-likelihood evaluated at the joint estimates is higher than that obtained from the uninformed models. This suggests that the additional constraint introduced by the non-decision time information helps guide the optimization procedure toward parameter regions that provide a better account of the behavioral data. In other words, uninformed models appear to converge to suboptimal parameter estimates, but the joint modeling framework helps regularize the estimation process and yields better estimates of the optimal parameters. We fully acknowledge that this use of BIC_{Behavior} for the joint

models departs from standard practice in the cognitive modeling literature. For this reason, we rely primarily on the joint likelihood-based model comparison as the formally valid criterion. The behavioral BIC is reported only to provide additional intuition regarding the goodness of fit on behavioral data and is used exclusively to compare NDT-informed models with their uninformed counterparts under identical evaluation criteria.

Also, to warn readers about this limitation, we included the following statement in the section where we defined BIC measures:

“It is worth noting that comparing joint and behavioral models using BIC_{Behavior} is uncommon and constitutes a limitation of the model comparison study.”

(4) Table 1 sets up the Study 1 comparisons, but there’s no row for the FT model. Similarly, Figures 10 and 13 would be more informative if they included FT predictions. This matters because, in Study 1, the FT model appears to fit aggregate accuracy better than the BIC-preferred collapsing model, currently shown only in Appendix 5. Some discussion of why would strengthen the argument.

In the revised manuscript, we have included the results for NDT-informed FT-DDM in the main text. However, we kept the FT-DDM with drift variability in the appendix, since none of the models in the main text include drift rate variability.

(5) In Figure 7, the degree of decay underestimation is obscured by using a density plot rather than a scatterplot, consistent with the other panels of the same figure. Presenting it the same way would make the mis-recovery more transparent. The accompanying text may also need clarification: when data are generated from an FT model with across-trial drift variability, the NDT-informed model seems to infer FT boundaries essentially. If that’s correct, the model must be misfitting the simulated data. This is actually a useful result as it suggests across-trial drift variability in FT models is discriminable from collapsing-threshold models. It would be good to make this explicit.

Regarding the visualization of decay rate estimation in Figure 7, we would like to clarify that, in the data-generating process for the FT model with across-trial drift variability, the decay rate is fixed at zero. That is, unlike the other parameters, there is only a single true value on the x-axis. If we were to present the decay rate recovery using a standard scatter plot (true vs. estimated values), all points would lie vertically above the single true value (zero), resulting in a vertical strip of points. While such a plot would technically be consistent with the other panels, it would not clearly convey the distributional properties of the estimated decay rates, specifically, which values are more likely under model misidentification. For this reason, we chose a density plot to more transparently illustrate the distribution of inferred decay rates when the true generating process is an FT model. We believe this representation more effectively communicates the extent and structure of mis-recovery. To avoid confusion, we have added a clarifying footnote in the revised manuscript explaining why a density plot was used in this specific panel.

Regarding distinguishing fixed-threshold models from collapsing-threshold models, as suggested in your comment 2, we conducted an additional simulation, and the results showed that when using the NDT-informed diffusion model, we can distinguish between both models with high accuracy. Specifically, we showed that incorporating the estimated decay rate value provided by the NDT-informed modeling approach can significantly improve the model recovery accuracy.

(6) Given the large recovery advantage of the exponential NDT-informed function over the hyperbolic one, the authors may want to consider whether the results favour adopting the former more generally. Given these findings, I would consider recommending the exponential NDT-informed model for future use.

Consistent with the reviewer's argument, we included the following text in the general discussion:

"Importantly, the exponential collapsing threshold exhibited substantially better parameter recovery and superior model recovery performance, suggesting that this specification may be preferable in future cognitive modeling applications."

(7) In Study 2 (Figure 13), all models qualitatively miss an interesting empirical pattern: under speed emphasis, errors are faster than corrects, while under accuracy emphasis, errors become slower. The error RT distribution in the speed condition is especially poorly captured. It would be helpful for the authors to comment, as it suggests that something theoretically relevant is missing from all models tested.

Thank you for mentioning this point. Because the models considered in the main text do not include across-trial variability in the starting point, the model cannot predict the fast error pattern observed in the speed condition of the second study. In the revised manuscript, we included a note on this point:

"The NDT-informed models' predictions are depicted in Figure 13. This figure shows that the FT-DDM overestimates the last RT quantiles for both correct and incorrect responses in both speed and accuracy conditions. However, the qualitative predictions of CT-DDMs align more closely with the empirical data. It is also worth noting that all the considered computational models misfit the incorrect responses in the speed condition. This misfit is to be linked to the presence of fast errors, specifically in the speed condition. Including the starting-point variability parameter in the model enables the model to predict fast errors. However, as the aim here was not merely to fit the data with the best possible model, but to test the NDT-informed modeling framework, we did not include starting-point variability in the model."

(8) The threshold visualisations extend to 3 seconds, yet both datasets show decisions mostly finishing by 1.5 seconds. Shortening the x-axis would better reflect the empirical RT distributions and avoid unintentionally overstating the timescale of the empirical decision processes.

In the new version of the manuscript, we shortened the x-axis in these plots. See Figures 9 and 12 in the new version of the manuscript.

Reviewer #1 (Recommendations for the authors):

(1) The manuscript should explicitly state how critical the log-normal assumption for NDT is, and whether there are caveats if it doesn't hold.

We agree with this suggestion. Therefore, we conducted an additional simulation study in which we assumed a normal distribution for non-decision time measurements and replicated the main results reported in simulation study 1 (see Appendix 8 in the new version of the manuscript). These simulation results reveal that independent of distributional assumptions on non-decision time measurements, constraining non-decision time can improve the estimation of the collapsing threshold.

(2) On page 14, one paragraph refers to five models and another to six. It is unclear which is correct.

We are sorry for the confusion. In the new version of the manuscript, we addressed this issue.

(3) In the Discussion: Hawkins & Heathcote (2021) found that NDT estimates in the TRDM recover well, but the timer-offset parameter does not (and hence is set to a fixed value). It would be interesting to test whether the NDT-informed approach could extend to that

parameter. The TRDM can also generate error RT distributions that are faster than correct RT distributions, which is the qualitative pattern that none of the present models capture in the speed condition of Study 2.

We appreciate the reviewer's suggestion, as it would be another important demonstration of the framework we built in the paper. Nevertheless, given the number of models already considered in the manuscript and the focus on collapsing threshold diffusion models, we believe that this addition would blur the focus of the present manuscript. Therefore, we have decided not to include TRDM in the manuscript. Nevertheless, we have addressed the reviewer's concern regarding the non-decision time estimation in the TRDM in the revised manuscript:

"Notably, this model can also predict faster error responses than the correct response, the pattern that is observed in the speed condition of Study 2. However, the authors reported poor parameter recovery for the onset of the timing process (Hawkins and Heathcote, 2021). Thus, the reliability issue here specifically concerns the estimation of the shift parameter of the timing accumulator. Informing the model with external estimates of non-decision time might, therefore, improve parameter recovery in this model as well."

Reviewer #2 (Public review):

Summary:

The authors use simulations and empirical data fitting in order to demonstrate that informing a decision model on estimates of single-trial non-decision time can guide the model to more reliable parameter estimates, especially when the model has collapsing bounds.

Strengths:

The paper is well written and motivated, with clear depth of knowledge in the areas of neurophysiology of decision-making, sequential sampling models, and, in particular, the phenomenon of collapsing decision bounds.

Two large-scale simulations are run to test parameter recovery, and two empirical datasets are fit and assessed; the fitting procedures themselves are state-of-the-art, and the study makes use of a very new and well-designed ERP decomposition algorithm that provides single-trial estimates of the duration of diffusion; the results provide inferences about the operation of decision bound collapse - all of this is impressive.

We appreciate your feedback and comments. Below, we provided a response for each comment.

Weaknesses:

(1) This is an interesting and promising idea, but a very important issue is not clear: it is an intuitive principle that information from an external empirical source can enhance the reliability of parameter estimates for a given model, but how can the overall BIC improve, unless it is in fact a different model? Unfortunately, it is not clear whether and how the model structure itself differs between the NDT-informed and non-NDT-informed cases. Ideally, they are the same actual model, but with one getting extra guidance on where to place the tau and/or sigma parameters from external measurements. The absence of sigma (non-decision time variance) estimates for the non-NDT-informed model, however, suggests it is different in structure, not just in its lack of constraints. If they were the same model, whether they do or do not possess non-decision time variability (which is not currently clear), the only possible reason that the NDT-informed model could achieve better BIC is because the non-NDT-informed model gets lost in the

fitting procedure and fails to find the global optimum. If they are in fact different models - for example, if the NDT-informed model is endowed with NDT variability, while the non-NDT-informed model is not - then the fit superiority doesn't necessarily say anything about an NDT-informed reliability boost, but rather just that a model with NDT variability fits better than one without.

To respond to this comment, we would like to note that the structural difference between NDT-informed and uninformed models is the assumption about non-decision time. In principle, the behavioral parts (i.e., parameters related to the diffusion part) of both NDT-informed and uninformed models are identical. However, during the estimation, the non-decision time in the NDT-informed model is subject to an additional constraint imposed by the neural data. In other words, in the uninformed model, non-decision time is estimated using the behavioral data by maximizing the likelihood of a CT-DDM. However, the NDT-informed model incorporates an additional data type, resulting in a different likelihood function (see Equation (3)). Specifically, in the NDT-informed model, we make an additional assumption regarding the non-decision time: it is set to the mean of the trial-level non-decision time measurements distribution derived from neural data (Equation (3) specifies the joint model structure). This additional assumption constrains the search space of the non-decision time parameter in the NDT-informed model. As in the main text, we assumed that non-decision time measurements obtained from neural data are log-normally distributed, where sigma is the shape parameter of the log-normal distribution. Therefore, sigma can represent the variability of the non-decision time measurements, and it is not the trial-to-trial non-decision time variability parameter. Indeed, the non-decision time parameter is fixed across trials in both models. Also, it should be clear from Equation (3) that sigma only appears in the second term of the joint likelihood, which corresponds to non-decision time measurements and not the CT-DDM term. Therefore, the behavioral parts of the NDT-informed models and Uninformed models are identical, and the NDT-informed models include one additional parameter (i.e., sigma) corresponding to the variability in non-decision time measurements.

Also, to explain how constraining non-decision time improves the BIC, we would like to clarify that constraining non-decision time using an additional data source constrains the search space for non-decision time, thereby leading to better parameter identification and, consequently, a better fit to behavioral data. We included the following text in the discussion section to make this explicit.

“This improvement likely reflects more accurate parameter estimation enabled by the additional information. In other words, constraining the non-decision time using neural measurements led the optimizer to estimate the CT-DDM parameter more accurately and, as a result, improve the fit to empirical data.”

(2) One reason this is unclear is that Footnote 4 says that this study did not allow trial-to-trial variability in nondecision time, but the entire premise of using variable external single-trial estimates of nondecision times (illustrated in Figure 2) assumes there is nondecision time variability and that we have access to its distribution.

We are sorry for the confusion. To respond to this comment, we would like to highlight that this modelling approach does not include any mechanism for across-trial variability in the behavioral part, as the likelihood of the choice behavior (see Equation (3)) does not include any variability parameter. As mentioned before, sigma represents the variability in non-decision measurements. However, the model does not propagate the across-trial variability of the neural data on the behavior side (unlike the models in Ghaderi-Kangavari et al. 2023). In other words, in this approach, none of the diffusion model's parameters include across-trial variability, and we have only considered and estimated neural data variability in the NDT-informed model. Also, to improve the manuscript's coherence, we removed this footnote.

(3) It is good that there is an Intro section to explain how the tradeoff between NDT and collapsing bound parameters renders them difficult to simultaneously identify, but I think it needs more work to make it clear. First of all, it is not impossible to identify both, in the same way as, say, pre- and postdecisional nondecision time components cannot be resolved from behaviour alone - the intro had already talked about how collapsing bounds impact RT distribution shapes in specific ways, and obviously mean (or invariant) NDT can't do that - it can only translate the whole distribution earlier/later on the time axis. This is at odds with the phrasing "one CANNOT estimate these three parameters simultaneously." So it should be first clarified that this tradeoff is not absolute. Second, many readers will wonder if it is simply a matter of characterising the bound collapse time course as beginning at accumulation onset, instead of stimulus offset - does that not sidestep the issue? Third, assuming the above can be explained, and there is a reason to keep the collapse function aligned to stimulus onset, could the tradeoff be illustrated by picking two distinct sets of parameter values for non-decision time, starting threshold, and decay rate, which produce almost identical bound dynamics as a function of RT? It is not going to work for most readers to simply give the formula on line 211 and say "There is a tradeoff." Most readers will need more hand-holding.

We are grateful for this comment. In response to this comment, which was also partly mentioned by the first reviewer, we first highlight that, in the presence of a nonlinear collapsing threshold, the effect of non-decision time is no longer linear, as it forces the threshold to take a specific value at the final stopping point. To illustrate the tradeoff between imprecise non-decision time estimation and collapsing threshold estimation, we conducted a simulation study. The results for the simulation study are presented in Appendix 1. In this study, we contaminated the non-decision time with noise at six levels (i.e., 0%, 2%, 4%, 6%, 8%, and 10%). The simulation results showed that increasing the noise level in non-decision time worsens the estimation of the starting threshold and decay rate parameters, indicating a trade-off between non-decision time and collapsing threshold parameters.

Regarding the second point, we would like to clarify that in this paper, consistent with other works on collapsing threshold, we assumed that the collapsing starts with evidence accumulation and not with stimulus onset, as the decision makers need a short amount of time for perceiving and encoding the stimulus (i.e., perceptual encoding time). Fixing the collapsing onset to the stimulus onset introduces an ad hoc assumption into the model (that the encoding time is zero), which is cognitively implausible. Therefore, although this assumption can mathematically resolve the issue, it is not cognitively plausible. However, one important point we did not consider in the previous version is the two-stage accumulation process models in which the collapsing onset occurs later than the evidence-accumulation onset. We have discussed the estimation of such models as a limitation in the general discussion:

"The collapsing threshold dynamics considered in this work (e.g., exponential and hyperbolic) impose a monotonically decreasing threshold over time. Although these dynamics are theoretically well motivated (Fudenberg et al., 2018; Frazier and Yu, 2007) and have been employed in several previous studies (e.g., Olschewski et al., 2025; Milosavljevic et al., 2010; Voskuilen et al., 2016), some research has proposed delayed-collapsing threshold models (e.g., Diederich and Oswald, 2016), in which the onset of threshold collapse does not coincide with the onset of evidence accumulation. Estimating such delayed-collapsing models may require more than simply constraining non-decision time, as the onset of threshold collapse must also be identified. A promising approach for addressing this challenge is the HMP method, which may provide additional temporal information about distinct cognitive processing stages. In particular, HMP may allow the onset of threshold collapse to be estimated as a separate cognitive stage. Future research should therefore investigate the

estimation of multi-stage evidence accumulation models (e.g., Diederich and Oswald, 2016; Diederich and Colonius, 2021) within the HMP framework.”

(4) A lognormal distribution is used as line 231 says it “must” produce a right-skew. Why? It is unusual for non-decision time distribution to be asymmetric in diffusion modeling, so this “must” statement must be fully explained and justified. Would I be right in saying that if either fixed or symmetrically distributed nondecision times were assumed, as in the majority of diffusion models, then the non-identifiability problem goes away? If the issue is one faced only by a special class of DDMs with lognormal NDT, this should be stated upfront.

We would like to clarify that this assumption is about the non-decision time measurements and not about the across-trial variability parameter of non-decision time. Although for computational convenience, non-decision time is often assumed to follow a normal distribution in diffusion modelling literature, some empirical studies have shown that the perceptual encoding time and total approximated non-decision time follow a right-skewed distribution. Moreover, as HMP estimations of non-decision time are usually right-skewed, we formalized the model with a log-normal distribution. However, we would like to note that the identifiability issue in collapsing-threshold diffusion models is not related to the distributional assumption over the non-decision time measurements, as poor parameter recovery of the collapsing threshold was also reported in Evans et al. (2020). To show that the distributional assumption about the non-decision time measurements does not affect the results, we conducted an additional simulation study in which we assumed a normal distribution for non-decision time measurements and showed that constraining non-decision time improves parameter recovery for collapsing threshold parameters. The results for this simulation are presented in Appendix 8 in the new version of the manuscript. We also revised line 231 as follows (see line 236 in the new version):

“Empirical studies on non-decision time measurement usually have reported a right-skewed distribution for their measurements (e.g., Weindel, 2021; Weindel et al., 2025). For instance, the measured perceptual encoding time and motor execution time reported by Weindel et al. (2025) are right-skewed. Therefore, we assume that the observed non-decision time measurements Z_n follow an approximate log-normal distribution, which is right-skewed (we will discuss how this distributional assumption can affect the results later). Thus, we model the non-decision time measurements Z_n , using a log-normal distribution with parameters μ and σ_z . The available measurements (i.e., observed data) at trial n can be represented as follows:”

(5) In the simulation study methods, is the only difference between NDT-informed and non-informed models that the non-NDT-informed must also estimate tau and sigma, whereas the NDT-informed model “knows” these two parameters and so only has the other three to estimate? And is it the exact same data that the two models are fit to, in each of the simulation runs? Why is sigma missing from the uninformed part of Figure 4? If it is nondecision time variability, shouldn’t the model at least be aware of the existence of sigma and try to estimate it, in order for this to be a meaningful comparison?

As mentioned in the response to your first comment, the difference between NDT-informed and uninformed models is the access to an additional source of data related to non-decision time, and sigma represents the shape parameter of the non-decision time measurements distribution. Therefore, sigma belongs only to the NDT-informed model, and, as in the uninformed model, there is no additional data source, so the model does not include sigma.

(6) I am curious to know whether a linear bound collapse suffers from the same identifiability issues with NDT, or was it not considered here because it is so suboptimal next to the hyperbolic/exponential?

Thank you very much for this comment. The main reason we did not include linear collapsing threshold models was the assessment by Evans et al. (2020), which indicated that, with sufficient trials, these models can be estimated reasonably well. However, to investigate whether constraining non-decision time can also improve the estimation of linear collapsing threshold models, we conducted an additional simulation study, which is reported in Appendix 3 in the new version of the manuscript. The simulation results confirm those reported by Evans et al. (2020) and show that the parameters of the uninformed linear models can be identified using more than 500 trials. Constraining the non-decision time using the NDT-informed diffusion modeling framework still improves parameter estimation in the linear collapsing threshold model and reduces the required number of trials for reliable estimation to 250. Especially, the estimation of the starting threshold improves significantly.

(7) The approach using HMP rests on the assumption that accumulation onset is marked by the peak of a certain neural event, but even if it is highly predictive of accumulation onset, depending on what it reflects, it could come systematically earlier or later than the actual accumulation onset. Could the authors comment on what implications this might have for the approach?

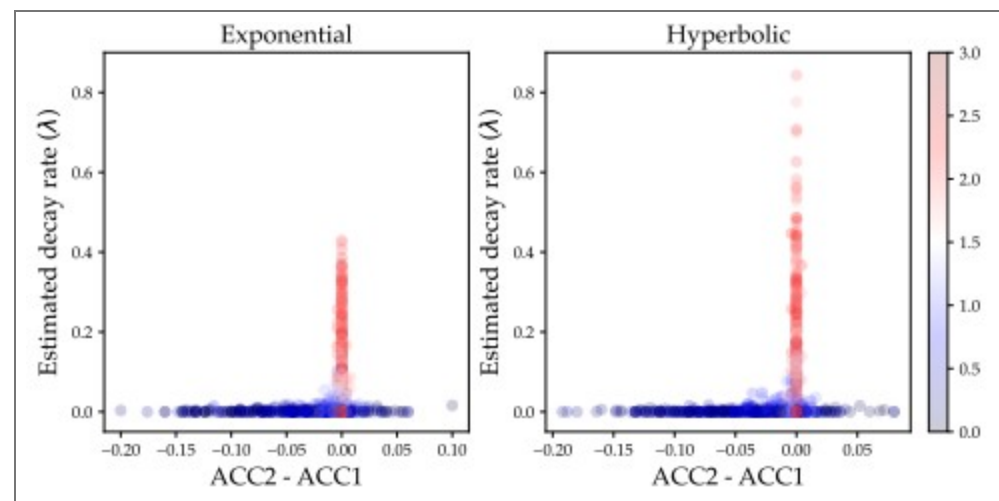
Thank you for mentioning this point. We first would like to point out that Weindel et al. (2024) showed that HMP can predict the underlying generative distribution of cognitive states with high precision and without systematic bias. Second, it is worth highlighting the results reported in Appendix 5 (i.e., “Bias in non-decision time measurements”). In this appendix, we discussed how bias in non-decision time measurements (i.e., systematic underestimation or overestimation) affects the estimation of the collapsing threshold. Particularly, see Figure 3 in Appendix 5. To make these results clearer in the main text, we included the following paragraph at the end of the results section in simulation study 1:

“Additionally, we examined the effect of systematic bias in non-decision time measurement on parameter recovery. Appendix 5 presents the parameter recovery simulation results in the presence of biased non-decision time measurement (i.e., systematically underestimated or overestimated). The results revealed that, even in the presence of biased non-decision time measurement, the actual generating parameters show a high correlation with the estimated parameters. Underestimation in non-decision time leads to overestimation in the starting threshold and decay rate. Conversely, overestimation in non-decision time leads to underestimation in both the starting threshold and non-decision time.”

(8) Figure 7: for this simulation, it would be helpful to know the degree to which you can get away with not equipping the model to capture drift rate variability, when the degree of that d.r. variability actually produces appreciable slow error rates. The approach here is to sample uniformly from ranges of the parameters, but how many of these produce data that can be reasonably recognised as similar to human behaviour on typical perceptual decision tasks? The authors point out that only 5% of fits estimate an appreciable bound collapse but if there are only 10% of the parameter vectors that produce data in a typical RT range with typical error rates etc, and half of these produce an appreciable downturn in accuracy for slower RT, and all of the latter represent that 5%, then that's quite a different story. An easy fix would be to plot estimated decay as a scatter plot against the rate of decline of accuracy from the median RT to the slowest RT, to visualise the degree to which slow errors can be absorbed by the no-dr-var model without falsely estimating steep bound collapse. In general, I'm not so sure of the value of this section, since, in principle, there is no getting around the fact that if what is in truth a drift-variability source of slow errors is fit with a model that can only capture it with a collapsing bound, it will estimate a collapsing bound, or just fail to capture those slow errors.

Thank you for this comment. We would like to first note that the aim of this section is to illustrate that NDT-informed modeling enables us to distinguish between CT-DDM and FT-DDM with drift variability (i.e., the two competing models that are relatively hard to distinguish). Therefore, we changed the name of this section to “Simulation study 2: Model recovery”. In the new version of the manuscript, we included a cross-model fitting simulation and a model recovery simulation in this section. The estimated decay rates in the cross-model fitting study indicate that, when the underlying generating model is FT-DDM with drift variability, parameter estimation using the NDT-informed approach yields precise inference. This is due to the estimated decay rate, which is very close to zero. The model recovery results also confirmed that incorporating the decay rate value into the model inference improves precision.

Moreover, to address the concern regarding the slow-error pattern in the simulated data, we examined the relationship between the slow-fast accuracy difference and the estimated decay rate. Specifically, we computed the difference between the accuracy of responses with response times below the median (ACC1) and those above the median (ACC2), and plotted this difference against the estimated decay rate. Author response image 1 presents the resulting scatter plot, with color indicating the drift rate. As shown in Author response image 1, incorrect inferences about the decay rate primarily occur at high drift rates. This pattern emerges because high drift rates produce very fast responses, leaving little time for the threshold to meaningfully collapse. Consequently, the behavioral signatures of collapsing-threshold and fixed-threshold models become increasingly similar under high drift conditions.



Author response image 1. Illustration of the difference between the accuracies of responses with response time below the median and those with response time above the median against the estimated decay rate. Colour shows the drift rate value.

Reviewer #2 (Recommendations for the authors):

- (1) Abstract: improves fit to behaviour in what way? Reliability or absolute quant fit to behavior? I.e., is it just helping constrain it so it finds the global opt?
- (2) Line 72 - Are these “neuroimaging” studies? Perhaps use the broader term “neuroscience.”
- (3) Line 95 - It's important because it may confuse readers how something dynamic like a collapsing bound could be resolved with, say, fMRI.

(4) Line 85 – “greater variability”... in what?

(5) Line 87 -Revise to avoid misconstruing a time on task effect - e.g., “higher error rate for trials with longer RT” is more explicit.

(6) Line 89-94 - It’s not clear what findings are being referenced here.

(7) Line 103 - External biases such as priors and relative value?

(8) Line 175 - Clarify this applies to any ddm, not just ctddm.

(9) Line 268 - Explain what portion N200 accounts for.

(10) Line 327 - Is this equation supposed to be for Delta-X, as opposed to $X(t+\delta t)$? If you want $X(t+\delta t)$ on the LHS, then $X(t)$ must be added to the RHS.

We are very thankful for such a precise evaluation and constructive comments. We addressed all the comments raised by the reviewer in the revised manuscript.

Reviewer #3 (Public review):

Summary:

The current paper addresses an important issue in evidence accumulation models: many modelers implement flat decision boundaries because the collapsing alternatives are hard to reliably estimate. Here, using simulations, the authors demonstrate that parameter recovery can be drastically improved by providing the model with additional data (specifically, an EEG-informed estimate of nondecision time). Moreover, in two empirical datasets, it is shown that those EEG-informed models provide a better fit to the data. The method seems sound and promising and might inform future work on the debate regarding flat vs collapsing choice boundaries. As an evidence-accumulation enthusiast, I am quite excited about this work, although for a broader audience, the immediate applicability of this approach seems limited because it does require EEG data (i.e., limiting widespread use of the method or e.g., answering questions about individual differences that require a very large N).

We are very grateful for your positive evaluation and your comments.

Reviewer #3 (Recommendations for the authors):

This is a very decent study, very well written and properly executed. Most of my comments below are suggestions for the authors as to how to make the story more compelling.

(1) I think the authors can do more to explain why the NDT-informed models fit better to empirical data. If the NDT-estimates are equal to the ground truth, then isn’t it possible that such models win simply because they have one parameter less? However, this is not what the authors are claiming, though, on l.567 it says that the fit is better because parameters are better estimated. I think it should be possible to dissociate these two accounts.

Thank you for this important comment. For clarification, both NDT-informed and uninformed models estimate the non-decision time parameter, and neither treats it as equal to the ground truth. However, the difference between these two models is that, in the NDT-informed model, the non-decision time is subject to an additional constraint; therefore, the number of behavioral parameters (i.e., parameters of the diffusion part) is identical for both NDT-informed and uninformed collapsing threshold models. Although the number of behavioral parameters is identical in the NDT-informed model, one parameter is constrained,

thereby limiting the model's complexity and flexibility compared to uninformed models. Consequently, the improvement in the fit can only be attributed to better parameter estimation in the NDT-informed models, resulting from constraining the non-decision time using neural measurements. To clarify that in the manuscript, we included the following text in the revised version:

"This improvement likely reflects more accurate parameter estimation enabled by the additional information. In other words, constraining the non-decision time using neural measurements led the optimizer to estimate the CT-DDM parameter more accurately and, as a result, improve the fit to empirical data."

(2) Given that so much of the writing focuses on the conclusion that collapsing boundary models & flat models, it is very odd that there is no comparison to a flat boundary model reported in the text. Why is the model in Supplement 5 not just included in the main text (and in Table 1)? This would make it so much easier for the reader.

To address this comment and also the similar point raised by Reviewer 1, we included the NDT-informed FT-DDM in the results section and compared the FT-DDM with CT-DDMs with respect to BIC_{joint} . However, we retained the other FT-DDM in the appendix because this model includes across-trial variability in drift rate, whereas the models in the main text do not.

(3) I would have appreciated a bit more background about the importance of the number of trials per participant. Given that the proposed method requires collecting EEG data (which is time and labour-intensive) I wonder to what extent you get similar improvements in parameter recovery by collecting more data per participant (which is usually cheap). Put differently, I would appreciate an additional matrix in Figure 5 for vanilla models.

The new version of Figure 5 in the revised manuscript now contains the goodness of parameter recovery for the uninformed CT-DDMs for different numbers of trials. As the simulation results suggest, even with 1000 trials, the parameters of the uninformed CT-DDMs are still not reliably identifiable. Therefore, these results suggest that although increasing the number of trials can slightly improve the parameter recovery of the CT-DDMs, it cannot fully resolve the reliability issue in their parameter recovery.

In addition, it is worth clarifying that, although the paper focuses on extracting non-decision time from the EEG signal using the HMP method, as discussed in the general discussion, the NDT-informed approach can also be employed with purely behavioral methods.

(4) L116-117: minor detail: I don't think it's fair to write that it's an open question whether or not thresholds collapse. I think it's fair to say that this is hard to show, and that the conditions under which it appears are unclear; but saying that it's still unclear whether this occurs at all seems unfair with regard to previous work.

Thank you for mentioning this point. We revised the mentioned sentence as follows in the new version of the manuscript:

"This issue is particularly critical given that the conditions under which individuals adjust their decision thresholds during a single trial remain an open question."

(5) Figure 5: Minor detail: It would be useful to mention in the figure or caption how many trials underlie these simulations, which is now somewhat buried in the text.

As Figure 5 shows sensitivity to the number of trials, and the number of trials is explicitly mentioned in the figure, we assume that the reviewer intended Figure 6. In the revised

manuscript, we explicitly specify the number of trials for the results reported in Figure 6. We simulated 500 trials for each noise level in this graph.

(6) Figure 10 and similar figures: I tried to figure out how corrects and incorrects differ, but couldn't see the difference. Can the authors use something more colorblind friendly (hope I didn't give up on being anonymous here)?

We are so sorry for the inconvenience. In the revised manuscript, we used different symbols to distinguish correct and incorrect data.

<https://doi.org/10.7554/eLife.109730.2.sa0>