# The Perils of Omitting Omissions when Modeling Evidence Accumulation

**Xiamin Leng*** **(xiamin_leng@brown.edu)**

**Alexander Fengler*** **(alexander_fengler@brown.edu)**

**Amitai Shenhav#** **(amitai_shenhav@brown.edu)**

**Michael J. Frank#** **(michael_frank@brown.edu)**

Department of Cognitive, Linguistic, and Psychological Sciences,
Carney Institute for Brain Science,
Brown University, Providence, RI 02912 USA

∗: These authors have contributed equally to this work.
#: These senior authors have contributed equally to this work.

**Abstract**

Choice deadlines are commonly imposed in decision-making research to incentivize speedy responses and sustained attention to the task settings. However, computational models of choice and response times routinely overlook this deadline, instead simply omitting trials past the deadline from further analysis. This choice is made under the implicit assumption that parameter estimation is not significantly affected by ignoring these omissions. Using new tools from likelihood-free inference, here we elucidate the degree to which omitting omissions, even in seemingly benign settings, can lead researchers astray. We explore the phenomenon using a Sequential Sampling Model (SSM) with collapsing boundaries as a test-bed.

**Keywords:** likelihood-free; Bayesian; DDM; sequential sampling models; omissions

## Introduction

Joint modeling of choices and response times is a core methodological staple of cognitive science. The broad class of evidence accumulation or sequential sampling models (SSMs), with the drift diffusion model (DDM) representing a widely used variant, form the dominant modeling paradigm for this purpose. SSMs, a very flexible class of models, grew out of the desire to capture detailed aspects of choices and RT distributions in increasingly complex experimental settings, which tested the limits of the original DDM (Ratcliff, 1978; Ratcliff, Smith, Brown, & McKoon, 2016).

According to SSMs, choices and response times are jointly generated as a result of a stochastic evidence accumulation process. The evidence evolves over time and eventually reaches a prescribed lower or upper boundary (sometimes referred to as the decision threshold).

Alternative variants of these models abound (for example, leaky competing accumulation model, race models, models which include collapsing boundaries etc.) (Usher & McClelland, 2001; Reynolds & Rhodes, 2009; Krajbich, Lu, Camerer, & Rangel, 2012; Hawkins, Forstmann, Wagenmakers, Ratcliff, & Brown, 2015; Malhotra, Leslie, Ludwig, & Bogacz, 2018). However, many such variants are rarely used in practice, due to limitations in the affordances of the supporting software infrastructure (Voss & Voss, 2007; Wiecki, Sofer, & Frank, 2013; Ahn, Haines, & Zhang, 2017; Fengler, Bera, Pedersen, & Frank, 2022). A particular limitation is that many theoretically interesting models lack closed-form expressions for the likelihood of choice-RT pairs, impeding the ability to perform Bayesian parameter estimation without requiring prohibitively long and expensive computational simulations (Fengler, Govindarajan, Chen, & Frank, 2021).

Prior work in this area has expanded the suite of models that can be rigorously evaluated beyond the canonical DDM (Fengler et al., 2021; Boelts, Lueckmann, Gao, & Macke, 2022; Durkan, Bekasov, Murray, & Papamakarios, 2019). Using simulators as training data generators to learn likelihood functions and / or posteriors cracked this bottleneck, with software support built around these innovations emerging (Fengler et al., 2022; Tejero-Cantero et al., 2020).

Here, we focus on the consequences of even seemingly more benign choices that researchers employ when modeling real data-especially the choice to exclude trials in which choices are not observed, and how likelihood-free methods are able to remedy them. Specifically, we focus on experimental designs that involve response deadlines, a typical form of experimental manipulation to induce prioritization of speed in decision making and a means to ensure sustained attention. While deadlines incentivize speedy decision making, a notable byproduct is a percentage of *omission* trials in which the participant did not commit a decision before the deadline. It is common-place in empirical practice to ignore omissions and focus the data analysis on the trials with committed responses. The implicit assumption is that a few omitted trials will not significantly impact parameter estimation, and therefore downstream conclusions. We will show here that this assumption is invalid and can indeed be pernicious.

We use the framework of Likelihood Approximation Networks (LANs) (Fengler et al., 2021), to empirically inves-



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 

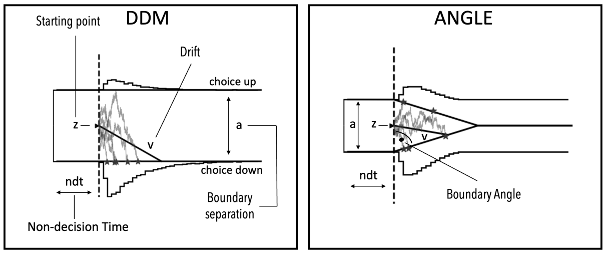


Figure 1: Graphic illustration of the standard DDM (left) and ANGLE (right) models. ANGLE model is used as the test-bed for our numerical experiments.

tigate the de-facto vulnerability towards misleading conclusions when *omitting omissions*. To account for omissions we add a second network that is trained directly on the omission probability $p(o|\theta, d)$, given model parameters $\theta$ and a deadline $d$ (Omission-Probability Network; OPN). We explain the resulting adjustments to likelihood computations in the methods section. With a series of numerical experiments, we show that parameter estimation (and therefore conclusions about mechanisms driving cross-condition differences in behavior) can be severely affected if omissions are not explicitly modeled, even when relatively few trials were omitted.

## Methods

### Cognitive Models

Because LANs, in conjunction with OPNs, allow us to investigate this phenomenon not only in the standard DDM model, but in a large class of SSMs, we focus our analysis here on an SSM with collapsing boundary (Cisek, Puskas, & El-Murr, 2009; Hawkins et al., 2015). This choice is motivated by two reasons: (i) to highlight that the risks we identify are not specialized toward the basic DDM, and (ii) because collapsing bounds have been used in particular for situations with response deadlines, as a rational decision maker can force themselves to decide with incomplete evidence (Frazier & Yu, 2007). As a test-bed for our investigation, we choose a DDM with linearly collapsing boundaries (ANGLE, right in Figure 1; standard DDM with fixed boundaries on left in Figure 1 for reference).

In general, SSMs are based on stochastic differential equations of the general form,

$$d\mathbf{X}_t = a(t,x)\,dt + b(t,x)\,d\mathbf{B}_t,\;\; \mathbf{X}_0 = z$$

where $\mathbf{X}$ represents the state of evidence in an accumulation process, $a(t,x)$ represents a *drift function*, which may depend on the position of $\mathbf{X}$, $b(t,x)$ represents a noise scaling process, $\mathbf{B}_t$ is the incremental noise process and $\mathbf{X}_0$ is the starting point. Decisions are made at the first point of exit of $\mathbf{X}_t$ from a region of interest, usually as soon as $\mathbf{X}$ exits the region prescribed by symmetric upper and lower bounds, prescribed as a function $f_{\text{bound}}$. This framework is very flexible, subsuming many theoretical models proposed in the literature (Usher & McClelland, 2001; Reynolds & Rhodes, 2009; Krajbich et al., 2012; Malhotra et al., 2018; Wieschen, Voss, & Radev, 2020).

The standard DDM is represented in this framework by setting $f^{DDM}_{\text{bound}}(t) = c$, a fixed boundary value, $a(t,x) = v$ a fixed drift over time, $b(t,x) = 1$, a fixed noise scaling over time, and setting the incremental noise process to be Gaussian. It is common to add a *non-decision time* $\tau$, to the model as a means to collect residual time spent on perceptual or motor processes not related to the choice process.

The ANGLE model, instead uses

$$f^{ANGLE}_{\text{bound}}(t; c, \theta) = a - \left( t * \frac{sin(\theta)}{cos(\theta)} \right)$$

.

as the shape of boundary and is otherwise identical to the DDM model. The linearly collapsing boundary serves to represent the concept of urgency in decision making, which has been proposed as theoretically relevant repeatedly (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Cisek et al., 2009; Malhotra et al., 2018). All of our analysis in the following focus on the ANGLE model.

### Likelihoods

The likelihoods for SSMs are described by a set of two defective distributions (one for each choice, $c = 1$ and $c = -1$), $f_c(t; \theta_{SSM})$, where,

$$\int_0^\infty f_{-1}(t; \theta_{SSM}) + f_1(t; \theta_{SSM})dt = 1$$

and $\theta_{SSM}$ represents the set of parameters for our given Sequential Sampling model. For our ANGLE model $\theta_{SSM} = (v, a, z, t, \theta)$, where $v$ is the *drift rate*, $a$ is the *boundary separation*, $z$ is the *starting-point bias*, $t$ is the *non-decision time* and $\theta$ is the angle of *boundary collapse*.

We are specifically concerned with incorporating *omissions*, however, and hence there are two types of data for which we need to specify likelihoods.

1. Standard: $d_i = (rt_i, c_i)$ where $rt_i \leq$ deadline, where both RT and choice are observed
2. Omission: $o_i = (rt_i, c_i)$ where $rt_i >$ deadline, where neither RT nor choice is recorded

For standard data-points we use our functions $f_c(rt; \theta_{SSM})$ above. For omissions, however, we need to compute the following integral,

$$p(\text{omission}|\theta_{SSM}, d) = \int_d^\infty f_{-1}(t; \theta_{SSM}) + f_1(t; \theta_{SSM})dt$$

In this integral, $d$ indicates the response deadline imposed by experiment and is treated as a known parameter. These integrals are not typically considered or available in relevant software packages for SSM parameter estimation, and instead researchers simply treat *omissions* as missing data, and thus

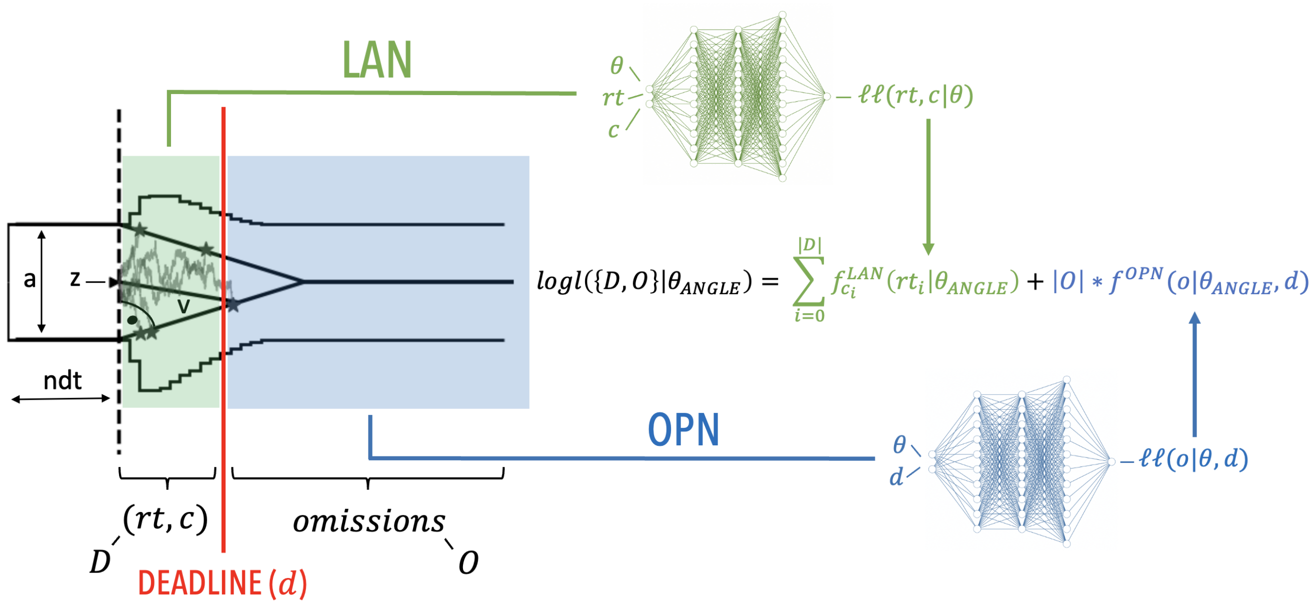


Figure 2: Illustration of the log-likelihood formulation with the help of a LAN and an OPN. The OPN provides a function, $f^{OPN}(.)$ for the log-likelihood of omissions, given a deadline setting. We simply multiply the log-likelihood of omissions by the number of omissed trials, which is enough for the numerical experiments introduced here. In principle, trial-wise deadlines can be incorporated at no extra cost. The LAN provides the function, $f_c^{LAN}(.)$ for the log-likelihood of observing a given $(rt, c)$ response time and choice pair prior to the deadline. Both networks take as parameters $\theta_{ANGLE}$ the underlying parameters of the ANGLE model. The LANs input adds to this the observed pair $(rt, c)$, while the OPN adds to this the *deadline* $d$. The general framework is applicable for a very large class of cognitive models, of which our ANGLE model is simply a concrete example (Fengler et al., 2021; Boelts et al., 2022; Fengler et al., 2022).

perform inference using only the probability density functions $f_c(t; \theta_{SSM})$.

Here we address this problem using the framework of likelihood approximation networks (LANs) (Fengler et al., 2021), to enable fast inference of the ANGLE model, while including the likelihood of omissions explicitly. In particular, we can use a simulator that returns the probability of *missing the deadline*, which can then be used to train a LAN along with its typical use to compute the functions $f_c(t; \theta_{SSM})$.

Hence we train two LANs for each model, one for the function $f_c^{LAN}(t|\theta_{SSM})$, and one for the function $f^{OPN}(\text{omission}|\theta_{SSM}, d)$. The log-likelihood of a given dataset $\{D, O\}$, is then computed as,

$$logl(\{D, O\}|\theta_{SSM}, d) = \sum_{i=0}^{|D|} f_{c_i}^{LAN}(rt_i|\theta_{SSM}) + |O| * f^{OPN}(\text{omission}|\theta_{SSM}, d)$$

this combined model enables estimation of $\theta_{SSM}$ via fitting observed choice data and omissions jointly.

## Numerical Experiments

To illustrate the extent to which misleading results may derive from data analysis that ignores omission trials, we use two sets of numerical experiments exemplified on ANGLE model. We note that in our numerical experiments, we use the simplest approach to omissions. Omissions strictly derive from the same process as recorded response times. In empirical data analysis, often a *lapse distribution* or *lapse probability* is introduced, adjusting the likelihood of outlier datapoints. Incorporating such lapse distributions may be fruitful for an even more comprehensive analysis in the future, however we would like to point out that the results we report in this paper happen for the *easiest possible* case. Adding lapse probabilities will only make the situation even more difficult, not easier.

**Parameter Recovery** First, we run a parameter recovery study for the ANGLE model. We simulate synthetic datasets, with ground-truth $v$, $z$, $t$ fixed ($v = 1.5, z = 0.5, t = 0.3$), and boundary parameters ($a$, $\theta$) sampled from a 2-D space of realistic values for each of the respective model parameters. We then impose a deadline (1.25s) and exclude parameter sets with too many omissions ($> 30\%$), so that the synthetic datasets generate omission percentages in the range of $0 - 30\%$. We then proceed with parameter inference in two ways:

1. Simply ignore omissions and evaluate only the LAN

(**LAN-only** model). This represents the workflow shortcut widely applied in the community.

2. Incorporate omissions via a second omission probability network (**LAN+OPN** model).

To reduce the influence of randomness in the simulated data, we repeat the parameter recovery for the same set of parameters on 20 different simulated datasets and evaluate parameter recovery based on the average across datasets.

**Synthetic Experiment** Second, we run four synthetic experiments in which we simulate data across two experimental conditions. In each experiment, the conditions share the same deadline (1.25s) and model parameters (e.g., v, a) except for the boundary collapse parameter θ. In Condition 1, the true boundary collapse is 0.9 rad (larger collapse), while in Condition 2 the collapse is 0.7 rad (smaller collapse) so the true difference in collapse is 0.2 rad.

For each experiment, we generated 1000 trials per condition (2000 trials in total) and obtained samples from the posterior distribution of Δθ (condition 1 - condition 2) by fitting **LAN-only** and **LAN+OPN** models with the simulated data. We then compared the posterior distribution with the ground truth Δθ. We evaluated the prediction of omission rate from the two types of models and compared the predictions with the rate in the simulated data.

## Results

**Parameter Recovery** When there is no omission in the simulated data, we observe similar performance on parameter recovery for the boundary parameters ($a$ and θ) (blue dots in Figure 3) regardless of whether the analysis is based on only a LAN or included an OPN. When omissions exist in the data (reddish dots in Figure 3), recovered values of $a$ and θ are higher than the true values when only LAN is used. This overestimation bias exists even when the rate of omission is very low ($< 5\%$). Specifically, we found that the LAN-only model overestimated $a$ especially when $a$ is high, and overestimated θ when θ is low. This leads to an overestimated change in $a$ (steeper lines in Figure 3A) and an underestimated change in θ (flattened lines in Figure 3B). These general patterns of bias in the LAN-only model persist across different levels of drift rate. Parameter recovery analysis also demonstrates that LAN-only model (but not LAN+OPN model) induces a magnified correlation between $a$ and θ so that higher $a$ is more likely to produce higher θ values even if the ground truth value of θ is the same (Figure 4).

As might be expected, the bias increases as omission rate increases (e.g., more missing data). However, it is worth noting that significant bias emerges at very low levels of omission already. In contrast, we obtained reliable parameter recovery with LAN+OPN model for various combinations of $(a, \theta)$ and different levels of omission rate. Altogether this suggests that omitting omissions is risky and may lead to misleading conclusions when parameter inference is an important aspect of a given study, even if omission rate is at a very low level.

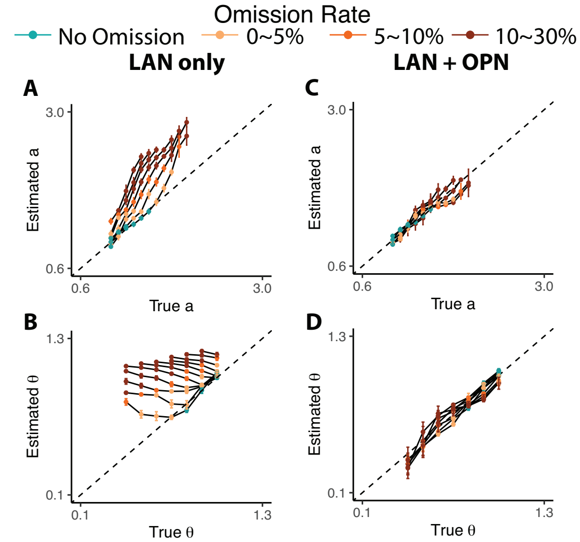


Figure 3: Parameter recovery performance for the ANGLE model, when either ignoring omissions (LAN-only model; **A, B**) or including them via the omission probability network as explained in Figure 2 (LAN+OPN model; **C, D**). **(A, C)** Patterns of parameter recovery from the two models for the *boundary separation* parameter $a$, varying true angle of the *boundary collapse*, θ (across solid lines). **(B, D)** Patterns for the *boundary collapse* parameter θ with varying true *boundary parameter a* (across solid lines). Error bars refer to 95% confidence interval of mean posteriors across multiple data sets from the same parameter configuration. Dashed lines represent identity lines.

**Synthetic Experiment** In addition to the overestimation bias, our parameter recovery analysis suggests that the LAN-only model may fail to correctly assign relative differences in parameters due to an experimental condition (here we focus on θ, but the analysis yields similar results with regards to the remaining model parameters). In the synthetic two-condition experiment, we found that the LAN-only model underestimates the difference in θ between two conditions ($\Delta\theta = 0.2$). while LAN+OPN model instead correctly recovers this difference between conditions Figure 5B. When the OPN was included, the model was able to recover the difference in collapse (θ) between the two synthetic experiment conditions for multiple combinations of drift rate and threshold (Figure 5B). Focusing on posterior predictives, the omission rates generated from the LAN-only model cannot capture the true omission rate of each condition (Figure 5C) while, again including the OPN resolves this discrepancy.

## Discussion

When implementing cognitive models, it is common to apply assumptions or shortcuts to simplify the computational pro-

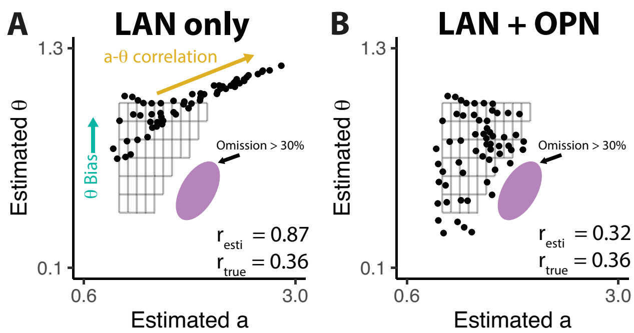


Figure 4: Estimation of θ is contaminated by both estimation biases and magnified *a*-θ correlation in LAN-only model **(A)** but not in LAN+OPN model **(B)**. Grids represent ground truth parameters. rs are Pearson correlations between *a* and θ.

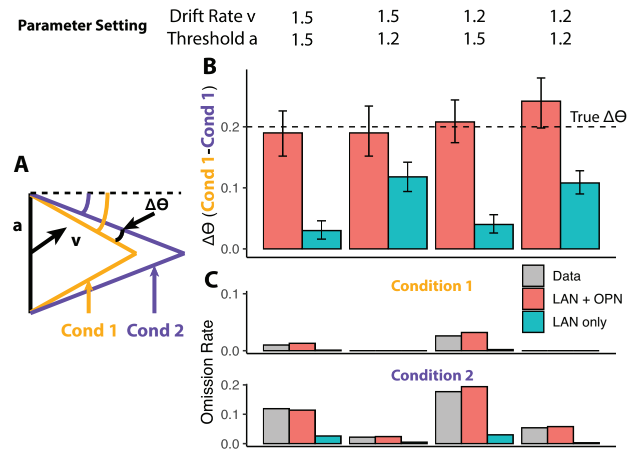


Figure 5: Results of synthetic experiments. **A)** Settings of experiments. Data is simulated from two conditions, with θ values of 0.9 and 0.7 respectively ($\Delta\theta = 0.2$), for a variety of settings of *drift rate* (v) and *threshold* (a) parameters. **B)** Recovery of collapse difference. While properly accounting for omissions correctly recovers Δθ, ignoring omissions severely underestimates the parameter difference across conditions. Error bars indicate 95% confidence interval of posterior distribution. **C)** Prediction of omission rate. Correspondingly omission rates in posterior predictive distributions are wrongly calibrated if omissions are not appropriately accounted for during the inference stage. Models fit without omissions systematically predict much lower overall omission rates and smaller variance in omission rate.

cedure. In the context of choice and RT modeling with deadlines, one such shortcut is to leave omissions out of the dataset when estimating computational model parameters. It seems straightforward to reason that, with a low percentage of omissions, parameter estimation will not be severely impacted by this methodological shortcut. Hence, this shortcut is inessential or benign. This shortcut allows experimental scientists to incorporate response deadlines in the experimental paradigms without concerns about omission trials. While this shortcut simplifies the cognitive model of choice behavior, its potential contaminating influence on modeling is ignored before thorough examination. Importantly, the lack of investigation in this shortcut originates from the fact that few modeling methods can properly account for omissions (Howard, Fox, Evans, Loft, & Houpt, 2023).

Enabled by modern computational tools (Fengler et al., 2021, 2022), in this paper, we sought to quantitatively examine the effects of this common shortcut and we come to the, perhaps surprising, following conclusion. As illustrated in Figures 3 even when omission rates are very low (5% or less), parameter recovery was severely impacted when ignoring omissions at inference. As we show via a synthetic experiment (Figure 5), this can result in highly misleading conclusions when comparing parameter values across separate experimental conditions, in search of mechanistic explanations of the effects of experimental manipulations.

We hope to have convincingly shown that omissions should not be disregarded, however small their number, when the goal of a study is parameter inference of computational cognitive models, especially (as is commonly the case) the comparison of inferred parameters across groups. On a broader scale, we think this investigation pointedly shows the value of continued re-examination of collective methodological choices. Taking advantage of the continuous development of novel inference tools, such efforts become progressively more feasible at the frontier of the computational sciences (Cranmer, Brehmer, & Louppe, 2020; Tejero-Cantero et al., 2020; Radev, Mertens, Voss, Ardizzone, & Köthe, 2020; Fengler et al., 2021).

The dearth of computational investigations to this effect (e.g. a rigorous workflow should employ a parameter recovery study for synthetic datasets otherwise equivalent to what is proposed for a given study, including trial or condition wise deadlines) stem from a simple phenomenon, which simultaneously explains the fields rigid focus on simple DDMs (Ratcliff et al., 2016) over an array of theoretically interesting model variations (Bogacz et al., 2006; Cisek et al., 2009; Wieschen et al., 2020) in the past: analytical convenience. Both cases-first the correct incorporation of deadlines in likelihood computations, as well as second, the exploration of model variants like e.g. the ANGLE model used as the basis for our investigation-imply a significant increase in effort as compared to the standard workflows established by canonical tools in the discipline (Wiecki et al., 2013; Voss & Voss, 2007). The effects of some of the resulting shortcuts remain untested. We wish to emphasize that the framework of simulation-based inference (Cranmer et al., 2020) and a focus on learning of likelihoods functions from simulations (Fengler et al., 2021; Boelts et al., 2022; Papamakarios, Sterratt, & Murray, 2019) in particular, present the emergence of a general purpose toolkit, which can fruitfully be employed for investigations like the present one on a much broader class of cognitive computational models as well as experimental designs. New toolboxes (Fengler et al., 2022; Tejero-Cantero et al., 2020) are designed to suggest a workflow that exposes the moving pieces in a computational modeling endeavor explicitly, while at the same time streamlining the otherwise analytically tedious aspects that adjustments of likelihood computations entail, through strong reliance on simulation. This approach fundamentally enabled the present investigation and is available to the wider community. We hope this investigation elucidates why it is not optional, but necessary, to use these modern tools to help diffuse computational best practices in the quest to avoid misleading conclusions from scientific work based on computational modeling.

## Acknowledgments

This work was supported by the Brainstorm Program at the Robert J. & Nancy D. Carney Institute for Brain Science and NIMH grants: P50MH119467 and P50MH106435-06A1.

## References

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hbayesdm package. *Computational Psychiatry (Cambridge, Mass.)*, *1*, 24.

Boelts, J., Lueckmann, J.-M., Gao, R., & Macke, J. H. (2022). Flexible and efficient simulation-based inference for models of decision-making. *Elife*, *11*, e77220.

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

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. In *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510–516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *Journal of Neuroscience*, *29*(37), 11560–11571.

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, *117*(48), 30055–30062.

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. (2019). Neural spline flows. *Advances in neural information processing systems*, *32*.

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill.

Fengler, A., Bera, K., Pedersen, M. L., & Frank, M. J. (2022). Beyond drift diffusion models: Fitting a broad class of decision and reinforcement learning models with hddm. *Journal of cognitive neuroscience*, *34*(10), 1780–1805.

Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (lans) for fast inference of simulation models in cognitive neuroscience. *Elife*, *10*, e65074.

Frazier, P., & Yu, A. J. (2007). Sequential hypothesis testing under stochastic deadlines. *Advances in neural information processing systems*, *20*.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, *35*(6), 2476–2484.

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year old. *Cognition and Brain Theory*, *6*, 287–317.

Howard, Z. L., Fox, E. L., Evans, N. J., Loft, S., & Houpt, J. (2023). An extension of the shifted wald model of human response times: Capturing the time dynamic properties of human cognition. *Psychonomic Bulletin & Review*, 1–21.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, *3*, 193.

Malhotra, G., Leslie, D. S., Ludwig, C. J., & Bogacz, R. (2018). Time-varying decision boundaries: insights from optimality analysis. *Psychonomic bulletin & review*, *25*(3), 971–996.

Matlock, T. (2001). *How real is fictive motion?* Doctoral dissertation, Psychology Department, University of California, Santa Cruz.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Ohlsson, S., & Langley, P. (1985). *Identifying solution paths in cognitive diagnosis* (Tech. Rep. No. CMU-RI-TR-85-2). Pittsburgh, PA: Carnegie Mellon University, The Robotics Institute.

Papamakarios, G., Sterratt, D., & Murray, I. (2019). Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In *The 22nd international conference on artificial intelligence and statistics* (pp. 837–848).

Radev, S. T., Mertens, U. K., 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*(4), 1452–1466.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, *20*(4), 260–281.

Reynolds, A. M., & Rhodes, C. J. (2009). The lévy flight paradigm: random search patterns and mechanisms. *Ecology*, *90*(4), 877–887.

Shrager, J., & Langley, P. (Eds.). (1990). *Computational models of scientific discovery and theory formation.* San Mateo, CA: Morgan Kaufmann.

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J., … Macke, J. H. (2020). Sbi–a toolkit for simulation-based inference. *arXiv preprint arXiv:2007.09114*.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, *108*(3), 550.

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior research methods*, *39*(4), 767–775.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). Hddm: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in neuroinformatics*, 14.

Wieschen, E. M., Voss, A., & Radev, S. (2020). Jumping to conclusion? a lévy flight model of decision making. *The Quantitative Methods for Psychology*, *16*(2), 120–132.