# Robustness of the Drift Diffusion Model Under Contaminated Trials

Rensselaer Polytechnic Institute

Harry Kim and Shichao Yang

April 26, 2025

# Introduction

This project investigates the robustness of the Drift Diffusion Model (DDM) in the presence of contaminated trials—those that arise from random guessing due to lapses in attention or distraction. While the DDM is a foundational cognitive model for binary decision-making and reaction times (RTs), it assumes all trials are driven by an underlying evidence accumulation process. Real-world data, however, often contain outlier trials from non-decision-based processes. This project explores how contamination affects DDM parameter recovery and inference, using both simulation-based analysis and full Bayesian inference.

# Methods

## Model Formulation and Assumptions

We utilize the standard DDM for binary decision-making, which models stochastic evidence accumulation toward decision boundaries. The key assumptions are:

- Evidence accumulation is continuous and stochastic.
- Decisions are binary.
- Non-decision time ($t_0$) accounts for fixed perceptual and motor delays.

Additionally, contamination is introduced by injecting non-DDM trials, representing random guessing with uniformly distributed RTs.

## Key Model Parameters

- Drift rate ($v$): the average evidence accumulation rate.
- Boundary separation ($a$): distance between decision thresholds.
- Non-decision time ($t_0$): fixed perceptual/motor delay.

In this project, noise ($\sigma$), representing within-trial variability, is held constant to avoid identifiability issues. Instead, we introduce trial-level noise by adding non-DDM trials. The goal is to determine whether the core DDM parameters can still be reliably estimated in the presence of such contamination.

## Priors and Full Model Specification

Prior predictive checks were performed to validate the choice of weakly informative priors:

- $v \sim \mathcal{N}(0, 1)$: normally distributed with mean 0 and standard deviation 1.
- $a \sim \text{HalfNormal}(1)$: a half-normal distribution centered at 1.
- $t_0 \sim \text{Uniform}(0.2, 0.5)$: uniformly distributed between 0.2 and 0.5 seconds.

Prior predictive simulations showed plausible distributions for RTs and choices, supporting these prior assumptions.

## Data and Simulation Setup

Each simulation condition includes 500 trials. Contamination levels are set to 0 percent, 10 percent, and 20 percent. For each trial:

- DDM trials are simulated using evidence accumulation until a boundary is crossed.
- Noise trials use a random choice (either 0 or 1) and a reaction time sampled from a uniform distribution between 0.3 and 2.5 seconds.

Each trial is labeled as either `ddm` or `noise`. Multiple simulations are performed for each setting to support reliable conclusions.

## Software and Libraries

Python 3.12, with NumPy, Pandas, Matplotlib, Seaborn, Stan, and ArviZ.

## Evaluation Metrics

- Prior predictive checks for parameter plausibility.
- Violin plots of RT distributions and bar plots of choice proportions.
- Quantitative summaries: Mean RT, RT standard deviation, and choice proportion across conditions.

## Bayesian Inference

Full Bayesian inference was implemented using Stan for the clean dataset, estimating:

- $v$, $a$, and $t_0$ with the priors specified above.

The Wiener first passage time distribution modeled the likelihood. Posterior samples were obtained via MCMC (No-U-Turn Sampler).

# Results

## Prior Predictive Checks

Sampling from the prior distributions produced reasonable RT and choice distributions, confirming that priors were appropriately weakly informative.

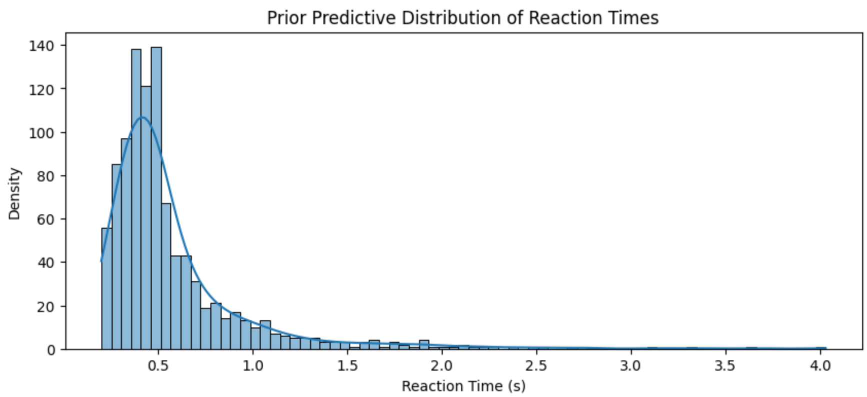


Figure 1: Prior predictive distribution of reaction times (RTs). Reaction times are centered within the plausible range (0.3–2.5 seconds).

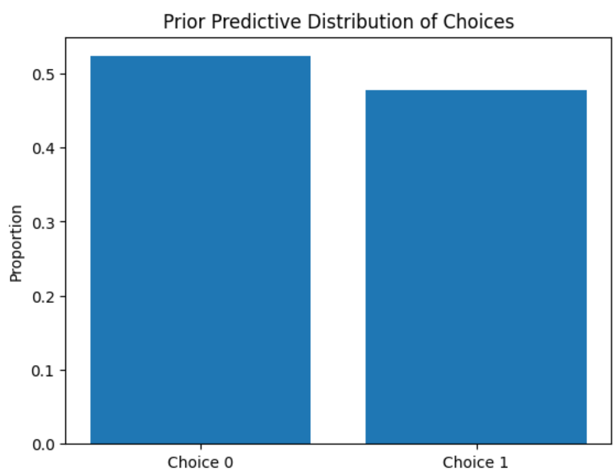


Figure 2: Prior predictive distribution of choices. The distribution shows balanced choice probabilities near 50%, reflecting weakly informative priors.

## Diagnostics

Bayesian inference was performed on the clean dataset using Stan. Convergence was assessed via trace plots and effective sample sizes. Posterior distributions were well-behaved, providing uncertainty quantification for $v$, $a$, and $t_0$.

## Interpretation

Contamination widened RT distributions and shifted choice proportions toward randomness (50%). Clean trials maintained structured decision dynamics with tighter RT clustering and choice biases aligned with drift.

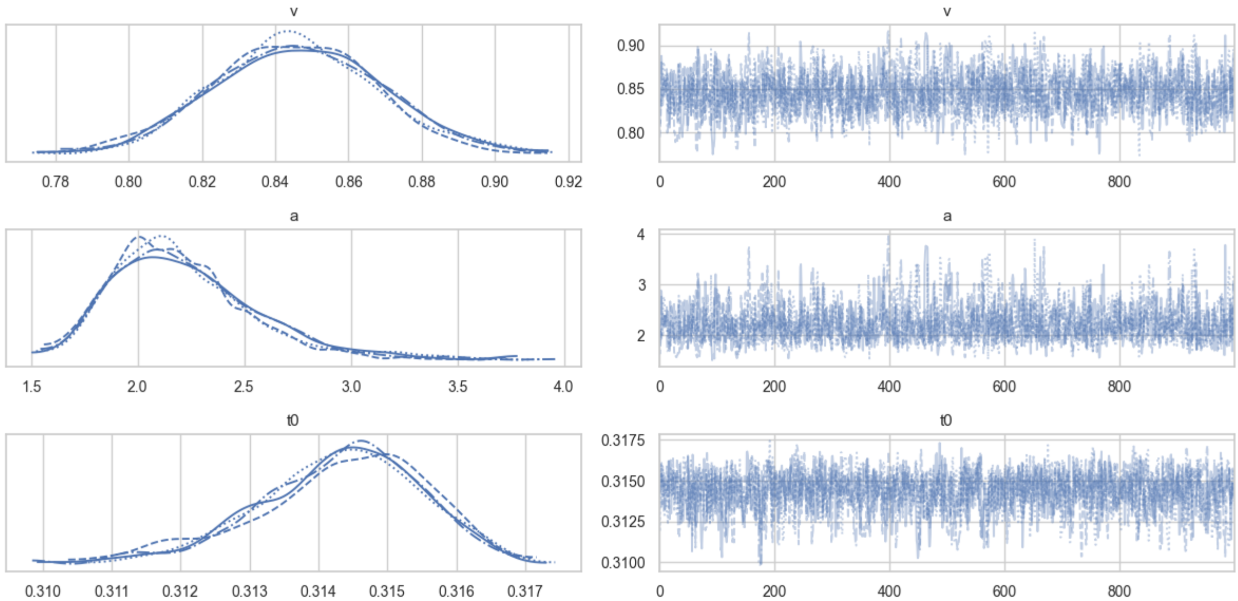


Figure 3: Posterior trace and density plots for drift rate ($v$), boundary separation ($a$), and non-decision time ($t_0$) estimated from the clean dataset. The left panels show posterior densities, indicating credible intervals and uncertainty, while the right panels display trace plots demonstrating proper mixing and convergence of MCMC chains.

## Validation

Ground truth parameters are known from simulation, which allows for qualitative validation. Clean conditions show clear decision dynamics, while noisy conditions flatten distributions and reduce observable effects of the model parameters.

## Sensitivity

The results show that DDM inference is sensitive to the inclusion of contaminated trials. Even small levels of noise (such as 20 percent) significantly degrade the clarity of the data. This makes parameter recovery more difficult and highlights the importance of identifying or modeling such noise during analysis.

## Quantifying Contamination Effects

Summary statistics revealed that contamination increased RT variability and pushed choice proportions toward 50%, indicative of random guessing:

- Mean RT increased in noisy conditions.
- RT variance broadened with contamination.
- Proportion of choice = 1 approached 0.5 in contaminated trials.

## Model Comparison

We compared:

- **Standard DDM (Bayesian)**: Fit to clean data, providing accurate parameter estimates.
- **Conceptual Mixture Model**: Proposed extension where trials are classified into DDM or noise components.

While the standard DDM struggled with contamination, the conceptual mixture model offers a pathway to explicitly model noise. Future work could implement this approach and compare models via WAIC or LOO.

## Bayesian Inference Results

Posterior distributions for $v$, $a$, and $t_0$ from clean data recovered parameters near their true values, with credible intervals reflecting uncertainty.

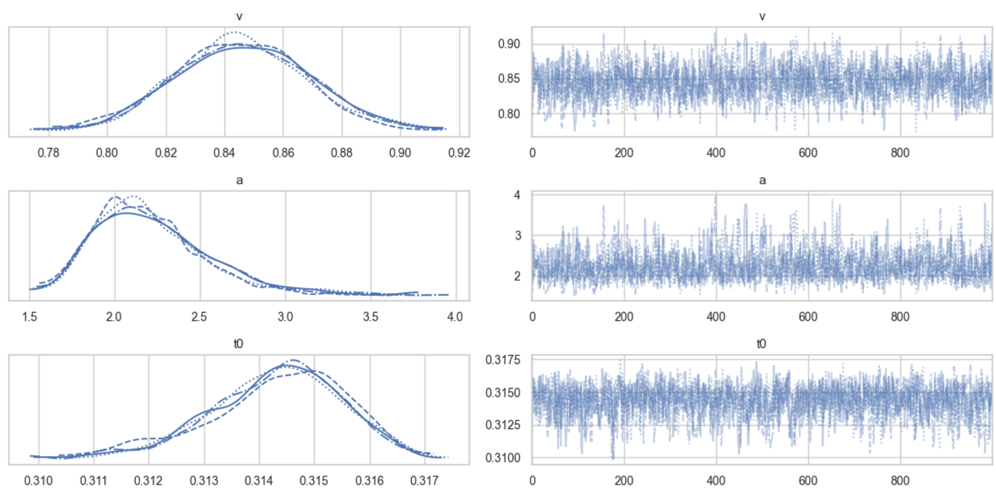


Figure 4: Trace plots and posterior densities for drift rate ($v$), boundary separation ($a$), and non-decision time ($t_0$). Chains mix well, and posterior distributions reflect credible intervals for parameter estimates.

# Discussion

This project demonstrates that even modest contamination degrades the DDM's inference quality. Key contributions include:

- Showing contamination effects through simulation and summary metrics.
- Implementing Bayesian inference for the standard DDM, providing posterior distributions for parameters.
- Conceptualizing a mixture model to address contamination.

**Future directions**:

- Implement a full mixture DDM with contamination modeling.
- Apply posterior predictive checks to assess generative adequacy.
- Compare models using WAIC or LOO.

This work highlights the need for robust cognitive models that account for real-world noise, improving inference reliability in decision-making studies.