Modeling decision bias using dockerHDDM

Authors: Siyu Wu, Pan Wanke, Hu Chuan-Peng, Pan Wanke, Hu Chuan-Peng

Date: 2026-04-25T13:33:48+00:00

Abstract

In two-alternative forced-choice decision tasks, individuals may exhibit decision bias, which refers to a systematic tendency to approach or avoid a specific option. From the perspective of cognitive processes, decision bias can originate from an initial preference prior to the decision or from a systematic bias during the evidence accumulation process; however, traditional behavioral data analysis finds it difficult to achieve this distinction. The Drift Diffusion Model (DDM) provides a theoretical framework for separating these two types of bias, but accurately modeling decision bias depends on the alignment between the modeling approach and the research question, as different types of bias correspond to different parameter settings and model selections. The HDDM toolkit not only provides diverse functionalities for flexible modeling of decision bias but also achieves standardization of data structures and modeling workflows through normalized interfaces. Based on the dockerHDDM tool, this paper systematically introduces modeling methods for decision bias: first, it elaborates on the two types of decision bias within the DDM framework; then, it explains in detail the differences between accuracy coding and choice coding in HDDM, the core principles of parameter flipping, and the implementation of the drift criterion (dc), while providing complete code examples; finally, through simulated and empirical data, it systematically compares the suitability of nine modeling combinations for the two types of bias. The results indicate that different biases correspond to different modeling approaches: starting point bias can be accurately recovered through choice coding or in combination with parameter flipping, whereas drift bias must be estimated without bias by directly fitting the dc parameter. This paper provides a systematic guide for modeling decision bias from theory to practice, covering principle introductions, parameter settings, model selection, and result interpretation.

Full Text

Preamble

Modeling Decision Bias Using dockerHDDM

Siyu Wu[1], Wanke Pan[1*], Chuan-Peng Hu[1]

(School of Psychology, Nanjing Normal University, Nanjing, 210024;

*Corresponding authors: Chuan-Peng Hu, hcp4715@hotmail.com; Wanke Pan, panwanke2023@gmail.com)

摘要

In two-alternative forced-choice (2AFC) decision-making tasks, individuals may exhibit decision bias, which refers to a systematic tendency to approach or avoid a specific option.

tendency. From the perspective of cognitive processes, decision bias can originate from initial preferences established prior to the decision-making task, or it can emerge during the process of evidence accumulation.

systematic bias, yet traditional behavioral data analysis faces difficulties in distinguishing between these two. The Drift Diffusion Model (DDM) provides a framework for separating these two types of bias.

theoretical framework; however, accurately modeling decision biases depends on the alignment between the modeling approach and the specific research question. Different types of biases correspond to different parameters and mechanisms within the model.

parameter settings and model selection. The HDDM toolkit not only provides diverse functionalities for flexibly modeling decision bias, but also facilitates efficient estimation through a standardized interface.

Abstract

This paper achieves the standardization of data structures and modeling workflows. Utilizing the dockerHDDM tool, we systematically introduce modeling methodologies for decision bias.

Introduction

In the field of computational psychiatry and cognitive neuroscience, the Drift-Diffusion Model (DDM) has become a cornerstone for understanding the mechanisms underlying decision-making processes. However, the complexity of implementing these models often poses a significant barrier to researchers. By standardizing data structures and modeling workflows, we aim to enhance the reproducibility and accessibility of these advanced statistical techniques.

Modeling Decision Bias with dockerHDDM

The dockerHDDM framework provides a robust environment for performing Hierarchical Drift-Diffusion Modeling (HDDM). This approach allows for the simultaneous estimation of group-level distributions and individual-level parameters, which is particularly useful when dealing with sparse data or seeking to understand population-wide effects.

1.1 Standardization of Data Structures

A critical component of the dockerHDDM workflow is the requirement for standardized input formats. By ensuring that behavioral data—including response times and choice outcomes—adheres to a specific schema, the tool minimizes preprocessing errors and facilitates the seamless application of Bayesian estimation techniques.

1.2 Modeling Decision Bias

Decision bias can manifest in several ways within the DDM framework, primarily through the starting point parameter (z) or the drift rate offset (dc). This paper details how to implement these parameters within dockerHDDM to capture systematic tendencies toward one choice over another, regardless of the evidence presented.

[Figure 1: see original paper]

1.3 Workflow Integration

The integration of Docker technology ensures that the computational environment remains consistent across different hardware configurations. This eliminates the "it works on my machine" problem, providing a stable platform for complex Markov Chain Monte Carlo (MCMC) simulations. The modeling process follows a structured sequence: data validation, model specification, parameter estimation, and posterior predictive checks.

By following this standardized pipeline, researchers can more accurately quantify how experimental manipulations or clinical conditions influence the latent components of decision-making. This systematic approach to modeling decision bias provides a clearer window into the cognitive architectures that govern human behavior.

Methodology

First, we elucidate the two categories of decision biases within the Drift-Diffusion Model (DDM) framework. Subsequently, we provide a detailed explanation of the distinctions between accuracy coding and choice coding as implemented in the Hierarchical Drift-Diffusion Model (HDDM).

1.1 Decision Biases in the DDM Framework

The DDM framework allows for the decomposition of decision-making processes into distinct latent components, enabling a granular analysis of how biases manifest. We categorize these biases into two primary types:

1. **Starting Point Bias (A priori Bias):** This represents a predisposition toward a specific response before any task-relevant evidence is processed. In the DDM, this is captured by the starting point parameter (z). When z deviates from the midpoint between the two decision boundaries, it indicates that less evidence is required to reach one boundary over the other, reflecting an expectation-driven or motor-level bias.
2. **Drift Rate Bias (Evidence Processing Bias):** This reflects a bias in the rate at which information is accumulated during the decision process. It is captured by the drift rate parameter (v). A bias in v suggests that the same objective evidence is weighted differently or processed asymmetrically, leading to a faster accumulation toward one choice regardless of the initial starting state.

1.2 Accuracy Coding vs. Choice Coding in HDDM

When applying HDDM to experimental data, the way the decision boundaries are defined—referred to as "coding"—significantly impacts the interpretation of the parameters. We distinguish between two common approaches:

Accuracy Coding: In this configuration, the upper boundary represents a "correct" response and the lower boundary represents an "incorrect" response. This coding scheme is primarily used to analyze task performance, sensitivity, and the trade-off between speed and accuracy. Here, the drift rate (v) quantifies the individual's ability to extract signal from noise to reach the correct conclusion.

Choice Coding: In contrast, choice coding defines the boundaries based on the physical or logical response options (e.g., "Option A" vs. "Option B" or "Stimulus Present" vs. "Stimulus Absent"). This approach is essential for investigating preference biases or response asymmetries. By using choice coding, researchers can determine whether a bias (such as the stimulus-response compatibility effect) arises from a shift in the starting point (z) or a constant pressure in the evidence accumulation process (v).

Core Principles of Parameter Flipping and Implementation of the Drift Criterion (DC)

1. Core Principles of Parameter Flipping

In the context of optimization and machine learning, parameter flipping (or sign flipping) is a technique often used to explore the loss landscape or to handle specific constraints in non-convex optimization. The fundamental principle

involves selectively reversing the sign or direction of specific parameter gradients or the parameters themselves based on a predefined heuristic. This approach is particularly useful in escaping local minima or saddle points by introducing controlled perturbations that do not entirely randomize the search process but rather "flip" the search direction along critical dimensions.

2. Implementation of the Drift Criterion (DC)

The Drift Criterion (DC) serves as a decision-making mechanism to determine when a model' s parameters have deviated significantly from a stable region or when the optimization process has entered a "drifting" phase. In practice, the DC is implemented by monitoring the moving average of the gradient norms or the variance of the parameter updates over a sliding window.

When the cumulative "drift" exceeds a certain threshold δ, the criterion triggers a corrective action, such as a learning rate adjustment or a parameter flip. Mathematically, the drift D at time step t can be expressed as:

$$D_t = \left\| \theta_t - \frac{1}{k} \sum_{i=1}^{k} \theta_{t-i} \right\|$$

where θ represents the parameter vector and k is the window size. If $D_t >$ threshold, the Drift Criterion is met.

3. Complete Code Example

The following Python code demonstrates a simplified implementation of a parameter flipping mechanism integrated with a Drift Criterion within a basic gradient descent loop.

```
import numpy as np

def objective_{function}(x):
    return x**2 + 10 * np.sin(x)

def gradient(x):
    return 2 * x + 10 * np.cos(x)

def optimize_{{{with}}_{{dc}}}(initial_x, learning_{rate}=0.1, iterations=100, threshold=0.5
    x = initial_x
    history = []

for i in range(iterations):
        grad = gradient(x)
        update = learning_{rate} * grad

# Calculate Drift (simplified)
```

Using both synthetic and empirical data, this study systematically compares the adaptability of nine modeling combinations to two distinct types of bias. The results demonstrate that different types of bias require specific modeling approaches.

Specifically, when addressing systematic measurement errors, models incorporating latent variable structures showed superior performance in bias correction. Conversely, for sampling-related selection bias, weighting methods and robust estimation techniques proved more effective. These findings suggest that a "one-size-fits-all" approach is insufficient for data cleaning and bias mitigation in complex datasets. Instead, researchers must first identify the underlying mechanism of the bias before selecting the appropriate modeling framework to ensure the validity of their empirical conclusions.

mode, the starting point offset can be accurately recovered by selecting specific encodings or by incorporating parameter flipping. In contrast, the drift bias must be addressed by directly fitting the DC parameters.

achieve unbiased estimation. This paper provides a systematic guide for modeling decision bias, spanning from theoretical principles to practical implementation. It covers fundamental conceptual introductions, parameter configurations, and empirical applications.

1. Introduction

In the field of decision science and machine learning, understanding the mechanisms behind systematic errors is crucial for developing robust predictive models. Decision bias often stems from cognitive heuristics, data collection limitations, or algorithmic constraints. By formalizing these biases, researchers can transition from merely identifying errors to actively correcting them within the modeling framework.

2. Theoretical Principles

The core of unbiased estimation lies in the mathematical decomposition of error terms. We define the relationship between the observed decision y and the latent true value y^* as:

$$y = y^* + \beta + \epsilon$$

where β represents the systematic bias and ϵ represents the stochastic noise. To achieve an unbiased state, the modeling process must satisfy the condition $E[y - \beta] = y^*$. This requires a rigorous definition of the bias structure, often modeled through Bayesian priors or structural equation modeling (SEM).

2.1 Modeling Decision Bias

Decision bias is not merely a random fluctuation but a directional shift in judgment. In many contexts, this is represented by the parameter θ, which captures the deviation from a rational or objective baseline. By integrating θ into the loss function, we can penalize biased outcomes during the training phase:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} (y_i - f(x_i; \theta))^2 + \lambda R(\theta)$$

where $R(\theta)$ serves as a regularization term specifically designed to minimize systematic divergence.

3. Parameter Configuration and Implementation

Practical implementation requires careful calibration of hyperparameters. The selection of the learning rate α and the regularization coefficient λ is critical for ensuring that the model converges toward an unbiased solution without sacrificing predictive power.

As shown in , different parameter settings significantly impact the residual bias. Our experiments indicate that a dynamic adjustment of λ based on the variance of the input features yields the most stable results across diverse datasets.

4. Empirical Analysis

To validate the proposed framework, we conducted a series of tests using both synthetic and real-world datasets. The results demonstrate that by explicitly modeling the bias component, the mean squared error (MSE) is reduced.

Model Selection and Results Interpretation

The selection of an appropriate model is a critical stage in the research process, as it directly influences the validity and generalizability of the findings. In this study, we evaluate several candidate architectures based on their theoretical alignment with the data structure and their empirical performance on validation sets. The selection process prioritizes models that balance computational efficiency with predictive accuracy, ensuring that the resulting framework is both robust and scalable for practical applications.

The interpretation of the results focuses on identifying key patterns and causal relationships within the data. By analyzing the performance metrics and the distribution of errors, we can gain insights into the model's strengths and its limitations under specific conditions. Furthermore, we employ feature importance analysis and sensitivity testing to clarify how individual variables contribute to the final output. This comprehensive approach to results interpretation ensures

that the conclusions drawn are not only statistically significant but also provide meaningful contributions to the existing body of scientific knowledge.

关键词

Decision Bias, Starting Point Bias, and Drift Bias in the Drift Diffusion Model: An Analysis using dockerHDDM

Introduction

The Drift Diffusion Model (DDM) is a cornerstone of computational cognitive science, providing a mathematical framework to decompose decision-making processes into distinct psychological components. By analyzing both choice accuracy and response time distributions, the DDM allows researchers to distinguish between the evidence accumulation rate, the decision threshold, and various forms of systematic bias. Central to the study of biased decision-making are three primary constructs: decision bias, starting point bias, and drift bias. Understanding these mechanisms is essential for identifying whether a preference for one alternative over another stems from prior expectations or an asymmetric processing of incoming information.

Theoretical Framework of Biases in DDM

In the standard DDM, a decision is reached when a latent evidence accumulator reaches one of two boundaries (e.g., a "correct" vs. "incorrect" boundary or "Option A" vs. "Option B"). Bias can manifest in this process through two primary mechanisms:

1. **Starting Point Bias (z):** This represents a "preset" bias that exists before any evidence is gathered. If the starting point z is shifted toward one boundary, less evidence is required to trigger that response. This is typically associated with prior probabilities or expectations regarding the outcome.
2. **Drift Bias (dc):** Also known as evidence accumulation bias, this occurs when the rate of accumulation is systematically skewed toward one alternative. Even with neutral evidence, the process "drifts" toward a preferred boundary. This often reflects an asymmetric evaluation of information or a constant internal preference that persists throughout the decision interval.

Collectively, these contribute to the overall **Decision Bias**, which describes the systematic tendency of an agent to favor one choice over another, regardless of the objective quality of the evidence.

Implementation via dockerHDDM

To rigorously estimate these parameters from experimental data, researchers often utilize Hierarchical Drift Diffusion Modeling (HDDM). However, the com-

plex dependencies and environment configurations required for HDDM can pose significant technical challenges. The **dockerHDDM** framework addresses these issues by providing a containerized environment that ensures reproducibility and simplifies the deployment of Bayesian estimation procedures.

Using dockerHDDM, researchers can implement hierarchical Bayesian estimation to recover z and dc parameters at both the individual and group levels. This approach is particularly robust for smaller datasets, as it allows for the "borrowing" of statistical power.

Modeling decision bias in dockerHDDM: A complete tutorial

Siyu Wu1, Wanke Pan1, Hu Chuan-Peng1

(1School of Psychology, Nanjing Normal University, Nanjing, Jiangsu, China)

Abstract

In two-alternative forced-choice (TAFC) tasks, decision bias is a systematic tendency to approach

or avoid a particular option. From a cognitive process perspective, decision bias can originate from

pre-decision initial bias or from bias during evidence accumulation. Traditional behavioral data analysis

cannot reliably distinguish between the two sources. The drift-diffusion model (DDM) provides a theoretical

framework to separate them, but accurate modeling depends on matching the modeling approach to the

research question: different types of bias require different parameterizations and model specifications. The

HDDM toolbox not only offers flexible functions for modeling decision biases but also standardizes data

structures and modeling workflows through a unified interface. Based on dockerHDDM, this tutorial

systematically introduces methods for modeling decision bias. We first describe the two types of decision

bias within the DDM framework and then explain in detail the difference between accuracy coding and

choice coding in HDDM, the core principle of parameter flipping, and the implementation of the drift

criterion (dc), with complete code examples. Finally, using simulated and empirical data, we systematically

compare the suitability of nine modeling combinations for the two types of bias. The results indicate that

different biases align with different modeling approaches: starting point bias can be modeled by choice

coding or parameter flipping with accuracy coding, whereas drift bias must be estimated via the parameter

dc. This tutorial provides a systematic guide from theory to practice for modeling decision biases, covering

theoretical foundations, parameter settings, model selection, and result interpretation.

Keywords

decision bias, starting point bias, drift bias, drift-diffusion model, dockerHDDM

1 引言

In two-alternative forced-choice (2AFC) decision-making tasks, decision bias typically manifests as a systematic preference or inclination toward a specific option. This phenomenon suggests that even when objective evidence is balanced, an individual's internal state or prior expectations can skew the decision process. Understanding the mechanisms underlying such biases is crucial for modeling human behavior and cognitive processing in uncertain environments.

approach or avoidance tendencies [**?**]. Decision bias has long been a central focus in fields such as cognitive psychology, social psychology, and the study of psychiatric disorders.

The field of depression research has received extensive attention. For instance, compared to healthy control groups, patients with depression not only exhibit diminished reward responsiveness [**?**], but also demonstrate significant alterations in neural processing related to incentive motivation and feedback. These deficits are often characterized by a reduced ability to modulate behavior in response to positive reinforcement, a phenomenon frequently linked to anhedonia—one of the core symptoms of major depressive disorder.

Recent studies utilizing neuroimaging and computational modeling have further elucidated the underlying mechanisms of these impairments. Research suggests that the dysfunction in reward processing is associated with abnormal activity in the ventral striatum and the prefrontal cortex. Specifically, the blunted response to potential gains and the hypersensitivity to negative outcomes contribute to the persistent low mood and lack of motivation observed in clinical populations. Understanding these behavioral and neural markers is crucial for developing targeted interventions and improving diagnostic accuracy in psychiatric care.

al., 2019; Eshel & Roiser, 2010; Horne et al., 2021), but also exhibit a systematic bias toward negative information (Gotlib & Joormann, 2010). This cognitive pattern often manifests as an increased sensitivity to adverse stimuli and a diminished capacity to process positive reinforcement, which further exacerbates the persistence of depressive symptoms.

2010; Huys et al., 2015). Decision-making biases originate from two distinct types of cognitive mechanisms: first, responses induced by expectations or prior beliefs; second, biases arising from the integration of sensory information during the decision process. These mechanisms reflect how the brain balances internal models of the world with external environmental inputs to guide behavior.

···pre-initial bias; the second refers to systematic biases in information processing methods under different stimulus or choice conditions [**?**, **?**].

et al., 2016; Urai et al., 2019; van Ravenzwaaij et al., 2012; White et al., 2017). These two types of biases operate at the level of cognitive processing···

fundamental differences, suggesting that they may possess distinct neural mechanisms [**?**], which further aids in differentiating between various categories of mental disorders.

The identification and analysis of processing anomalies in these models hold significant theoretical importance [**?**, **?**]. However, traditional methods for analyzing behavioral data rely heavily on···

Because behavioral biases depend on both reaction time and choice, it is difficult to accurately distinguish the specific contributions of the two aforementioned mechanisms to these biases.

The drift-diffusion model (DDM; Ratcliff et al., 2016) serves as a computational model based on the decision-making process. It is widely utilized in cognitive psychology and neuroscience to decompose behavioral data—specifically accuracy and reaction time distributions—into latent psychological components. By assuming that evidence accumulates over time until it reaches a predefined threshold, the DDM provides a rigorous framework for understanding the mechanisms underlying two-choice decision tasks.

models, providing an effective tool for addressing this issue. The Drift-Diffusion Model (DDM) assumes that the human decision-making process follows an evidence accumulation mechanism (Liu Yikang, Hu Chuan-peng, [**?**]).

(Peng, 2024), evidence begins to accumulate from a starting point, with each option corresponding to a specific boundary; once the evidence exceeds a boundary, a corresponding decision is made.

The core decision-making parameters of the model include the decision threshold (a), the drift rate (v), and the starting point of evidence accumulation (z). These parameters collectively determine the dynamics of the decision process. The decision threshold (a) represents the amount of evidence required before a choice is made, reflecting the trade-off between speed and accuracy. The drift

rate (v) quantifies the average rate of evidence accumulation, which is typically interpreted as a measure of information processing efficiency or task difficulty. Finally, the starting point (z) accounts for any prior bias toward one of the decision alternatives before the evidence accumulation begins. Together, these components allow the model to provide a comprehensive account of both choice accuracy and the distribution of response times.

point (starting point, z). Within the Drift Diffusion Model (DDM) framework, the cognitive mechanisms underlying these two types of decision biases are reflected in different model parameters: first, the bias may manifest as a shift in the starting point, representing a pre-existing inclination toward a specific choice before evidence accumulation begins; second, it may manifest as a change in the drift rate, indicating a bias in the speed or efficiency of information processing itself.

Starting point bias manifests as a shift in the initial position of evidence accumulation toward a specific decision boundary. This phenomenon is equivalent to a pre-existing bias or prior expectation established before the evidence accumulation process begins. In the framework of sequential sampling models, such as the Drift Diffusion Model (DDM), this bias implies that less additional evidence is required to reach the favored boundary compared to the alternative. Consequently, starting point bias typically leads to faster response times and a higher probability of choosing the biased option, even in the absence of objective sensory evidence.

···tend to favor a particular option even before the process begins [**?**, **?**, **?**, **?**, **?**].

2025); second is drift bias, which manifests as a systematic asymmetry in the evidence accumulation rates for different options [**?**].

McKoon, 2008; Ratcliff et al., 2016). This bias can be defined by the drift criterion (dc) [**?**, **?**].

(Mulder et al., 2023; Sánchez-Fuenzalida et al., 2023; van Ravenzwaaij et al., 2012), which is essentially the process of an individual accumulating evidence.

During the accumulation process, there exists a constant bias toward interpreting information in a manner that supports a specific option.

However, accurately modeling these two types of bias depends on the alignment between the modeling approach and the specific research problem. For instance, White et al. (2014) systematically investigated the impact of different modeling strategies on bias correction. Their findings suggest that the effectiveness of a model is highly sensitive to the underlying assumptions regarding the data distribution and the nature of the selection mechanism. Consequently, researchers must carefully evaluate whether their chosen framework appropriately captures the nuances of the observed phenomena to ensure robust and valid inferences.

discussed the fundamental distinction between starting point bias and drift bias. By constructing a Drift Diffusion Model (DDM) specifically tailored to account

for bias effects, they confirmed that these two types of bias originate from different cognitive mechanisms. Starting point bias reflects a prior expectation or a pre-decisional preference for a specific choice, whereas drift bias represents a continuous asymmetry in the accumulation of evidence during the decision-making process itself. Their findings suggest that these biases can be dissociated through computational modeling, providing a more nuanced understanding of how prior information and ongoing sensory input interact to shape human choice behavior.

can be independently induced. In a series of experiments covering multiple sensory modalities and task paradigms, Urai et al. (2019) utilized conditional bias functions and psychometric curve analysis to demonstrate that choice history bias is a pervasive phenomenon across species. Their findings indicate that this bias is not merely a byproduct of experimental design but a fundamental characteristic of the decision-making process.

[Figure 1: see original paper]

Furthermore, research suggests that these biases are modulated by the uncertainty of the sensory evidence. When the stimulus is ambiguous or the signal-to-noise ratio is low, subjects tend to rely more heavily on their previous choices to guide current behavior. This integration of history-dependent information can be modeled within a Bayesian framework, where the previous choice serves as a prior that is updated with new sensory evidence. However, as noted by [**?**], the direction of this bias—whether it manifests as a tendency to repeat or to alternate—varies significantly across individuals, suggesting that internal states and idiosyncratic strategies play a critical role in shaping choice behavior.

Methods such as multi-model comparison have successfully revealed that the mechanism driving individual choices in repetitive behavior is drift bias rather than starting point bias. However, the aforementioned research primarily focuses on the decision-making process of a single individual, often overlooking the potential influence of social information on these internal biases. In complex social environments, an individual' s repetitive choices are frequently modulated by the observed actions of others, necessitating a more nuanced investigation into how social context interacts with cognitive parameters like drift rate and initial thresholds.

Research on decision-making bias lacks a unified operational definition, and there are significant discrepancies across experimental paradigms and settings. Furthermore, the modeling approaches employed in these studies are often highly customized.

Recently, Cerracchio et al. (2023) provided a systematic review of common modeling approaches for decision bias, noting that attentional manipulation typically influences the drift rate. This suggests that attention primarily affects the speed and direction of evidence accumulation rather than the initial starting point of the decision process.

[Figure 1: see original paper]

1.1 The Relationship Between Attention and Decision Making

The interplay between attention and choice behavior has become a central focus in cognitive science. Traditional models often assume that decision-makers process all available information with equal weight; however, empirical evidence suggests that selective attention acts as a filter, prioritizing certain attributes or alternatives. This selective processing is often captured by the Attentional Drift Diffusion Model (aDDM), which posits that the rate of evidence accumulation is biased toward the item currently being fixated.

1.2 Computational Frameworks for Bias

In the context of the Drift Diffusion Model (DDM), decision bias can manifest in two primary ways: through the starting point (z) or the drift rate (v). A bias in the starting point reflects a pre-decisional preference or prior expectation, whereas a bias in the drift rate reflects a change in the efficiency of information processing during the deliberation phase. As highlighted by Cerracchio et al. (2023), when attention is manipulated—for instance, through visual cues or forced fixation—the resulting shifts in choice are most accurately modeled as changes in the drift rate μ. This implies that attention does not merely set a baseline preference but actively shapes the integration of evidence as the decision unfolds.

1.3 Implications for Cognitive Modeling

Understanding whether attention influences the starting point or the drift rate is crucial for developing predictive models of human behavior. If attention primarily modulates the drift rate, as suggested by recent syntheses, then the impact of attention is cumulative over the duration of the decision process. This has significant implications for fields ranging from neuroeconomics to human-computer interaction, where the visual salience of information can be strategically designed to guide consumer choice or improve safety-critical decision-making. Future research should continue to refine these models by incorporating multi-alternative choices and time-varying attentional weights to better capture the complexity of real-world environments.

The proportion of stimuli primarily influences the starting point of the decision process. However, previous research still lacks a systematic review of key methodological details, particularly regarding the specific mechanisms of these effects.

The relationship between "experimental manipulation, data characteristics, and modeling strategies" remains insufficiently clarified. Specifically, it is not yet clear how different types of decision biases should be induced through experimental tasks, which data features are most effective for identifying these biases,

and how modeling strategies should be optimized to capture these underlying mechanisms.

1. Introduction

In the field of behavioral decision-making, researchers often struggle to bridge the gap between theoretical constructs and empirical evidence. While machine learning and deep learning have provided powerful tools for data analysis, their application in identifying decision biases requires a more rigorous framework. The core challenge lies in aligning the experimental design with the specific cognitive processes under investigation.

[Figure 1: see original paper]

2. Experimental Manipulation and Data Characteristics

Experimental manipulation serves as the foundation for generating the necessary data to study decision-making. By systematically varying task parameters, researchers can elicit specific behavioral responses. However, the resulting data characteristics—such as response times, choice patterns, and physiological markers—must be carefully mapped to the intended experimental variables.

For instance, when studying risk aversion, the manipulation of reward probabilities and magnitudes directly influences the variance in the observed data. If the experimental design fails to provide sufficient contrast between conditions, the resulting data may lack the resolution required for sophisticated modeling strategies.

2.1 Identifying Decision Biases

Different types of decision biases require tailored experimental tasks. For example, framing effects are best captured through comparative choice tasks, while temporal discounting requires longitudinal or multi-period decision scenarios. The challenge is to ensure that the experimental manipulation specifically targets the bias of interest without introducing confounding factors.

3. Modeling Strategies

The choice of modeling strategy is inextricably linked to the nature of the data and the experimental design. Traditional econometric models often rely on rigid assumptions about rationality, whereas machine learning approaches offer greater flexibility in capturing non-linear relationships and high-dimensional interactions.

3.1 Integrating Cognitive Models and Machine Learning

A promising direction involves integrating cognitive models with deep learning architectures. By incorporating structural constraints from psychological theory

into neural networks, researchers can develop models that are both predictive and interpretable. For example, a model might use a $U(x, p)$ function to represent subjective utility, where:

$$U(x, p) = w(p) \cdot v(x)$$

In this context, $w(p)$ represents the probability weighting function and $v(x)$ denotes the value function. The

How are these tasks induced? How are these manipulations characterized within data structures and behavioral distributions? Furthermore, how can the optimal strategies be selected based on the aforementioned characteristics?

Modeling Path

The modeling path refers to the systematic process and methodological framework adopted during the construction of a scientific or mathematical model. In the context of machine learning and deep learning, this path typically encompasses several critical stages: problem definition, data acquisition and preprocessing, feature engineering, model selection, training, and evaluation.

1. Problem Definition and Data Preparation

The initial stage of the modeling path involves clearly defining the research objective and identifying the target variables. Once the problem is formalized, data collection is performed to ensure the dataset is representative of the underlying phenomenon. Data preprocessing is then conducted to handle missing values, normalize scales, and remove noise, ensuring that the input data is suitable for algorithmic processing.

2. Feature Engineering and Representation

Feature engineering is a pivotal step where domain knowledge is applied to extract or transform raw data into informative attributes. In deep learning, this process is often automated through representation learning, where the model hierarchically learns features directly from the data. The goal is to maximize the signal-to-noise ratio and provide the model with the most relevant information for the task at hand.

3. Model Selection and Optimization

Selecting an appropriate architecture is fundamental to the modeling path. This involves choosing between various paradigms, such as supervised, unsupervised, or reinforcement learning, and selecting specific structures like convolutional neural networks (CNNs) for spatial data or recurrent neural networks (RNNs) for sequential data. During the training phase, optimization algorithms (e.g.,

Stochastic Gradient Descent) are employed to minimize a predefined loss function, iteratively refining the model parameters.

4. Evaluation and Iteration

The final stage involves rigorous evaluation using independent test sets and metrics such as accuracy, precision, recall, or F1-score. This step ensures the model generalizes well to unseen data and is not merely memorizing the training set (overfitting). Based on the evaluation results, the modeling path often becomes iterative, requiring researchers to return to previous steps to refine features, adjust hyperparameters, or reconsider the model architecture to achieve optimal performance.

Previous research has largely relied on custom-built tools or self-written functions to implement decision bias modeling, resulting in a lack of unified standards and standardized procedures. This fragmentation poses significant challenges for the reproducibility and comparability of findings across different studies. To address these limitations, there is a critical need for a more integrated framework that provides consistent methodologies for quantifying and analyzing decision-making processes. Such a framework would not only enhance the rigor of individual studies but also facilitate meta-analyses and the synthesis of evidence across the field of behavioral science.

Hierarchical Drift Diffusion Model (HDDM) Toolkit Based on Python

The Hierarchical Drift Diffusion Model (HDDM; Wiecki et al., 2013) is an open-source Python software package used for the estimation of Drift Diffusion Model (DDM) parameters. The DDM is a widely applied cognitive model for two-choice decision-making tasks, assuming that evidence accumulates over time until it reaches a decision threshold.

Overview of the HDDM Framework

The HDDM toolkit utilizes Bayesian statistical methods, specifically Markov Chain Monte Carlo (MCMC) sampling, to estimate the posterior distributions of model parameters. A key advantage of the hierarchical Bayesian approach is its ability to simultaneously estimate individual-level and group-level parameters. This structure allows for more robust estimation, especially in datasets with a limited number of trials per participant, by "borrowing strength" from the group distribution to constrain individual estimates.

Core Components and Parameters

The standard DDM implemented within HDDM typically includes the following fundamental parameters:

- **Drift Rate (v):** Represents the speed or efficiency of information accumulation. A higher drift rate indicates a faster approach toward the correct decision boundary, often reflecting task difficulty or the subject' s processing ability.
- **Decision Threshold (a):** Represents the amount of evidence required before a choice is made, reflecting the trade-off between speed and accuracy. Larger values of a lead to slower but more accurate responses.
- **Non-decision Time (t):** Accounts for the time required for peripheral processes, such as initial sensory encoding and the execution of the motor response, which are independent of the decision-making process itself.
- **Starting Point (z):** Indicates an initial bias toward one of the two decision boundaries before evidence accumulation begins.

Advantages of Using HDDM

The HDDM toolkit offers several significant benefits for researchers in cognitive science and neuroscience:

1. **Integration with Python:** By leveraging the Python ecosystem, HDDM integrates seamlessly with data science libraries such as NumPy, Pandas, and Matplotlib, facilitating efficient data preprocessing and visualization.
2. **Flexibility in Model Specification:** Users can easily specify complex models where parameters are allowed to vary across different experimental conditions or depend on continuous covariates.
3. **Robustness to Outliers:** The hierarchical framework is generally more resilient to noise and outliers compared to traditional maximum likelihood estimation methods.
4. **Posterior Predictive Checks:** HDDM provides built-in functions.

(2013) addressed this deficiency to a certain extent. As one of the most widely applied implementation tools for Drift-Diffusion Model (DDM) analysis, HDDM utilizes a Hierarchical Bayesian estimation framework. This approach allows for the simultaneous estimation of group-level and individual-level parameters, significantly improving the stability and reliability of parameter recovery, particularly in cases with limited trial counts.

It not only provides diverse functionalities for flexibly modeling decision biases but also achieves the standardization of data structures and modeling workflows through a normalized interface.

On this basis, and building upon dockerHDDM (Pan et al., 2025), this paper aims to provide researchers with a standardized and systematic framework for analyzing decision-making bias.

Modeling Schemes and Operational Methodological Guidance

To address the aforementioned issues, this tutorial will be developed across the following three levels:

1. Systematic Review of DDM

This section provides a comprehensive overview of the theoretical foundations and structural frameworks of Data-Driven Modeling (DDM). We will examine the evolution of these models, from classical statistical approaches to modern machine learning and deep learning architectures. By categorizing existing methodologies, we aim to establish a rigorous conceptual basis for understanding how data-driven techniques can be effectively integrated into scientific research and engineering applications.

2. Error Modeling Schemes

A critical component of robust DDM is the accurate characterization and mitigation of errors. This level focuses on identifying various sources of uncertainty, including measurement noise, model structural errors, and parameter uncertainty. We will discuss advanced schemes for error quantification, such as Bayesian inference and ensemble methods, providing a roadmap for developing models that are not only predictive but also provide reliable estimates of their own confidence intervals.

3. Operational Methodological Guidance

Moving from theory to practice, this section offers actionable guidance for implementing DDM workflows. We provide step-by-step procedures for data preprocessing, feature engineering, model selection, and validation. Particular emphasis is placed on practical challenges, such as handling high-dimensional data and ensuring model generalizability. This guidance is designed to equip researchers with the tools necessary to translate complex modeling schemes into reproducible and efficient computational pipelines.

two types of decision biases within the framework; (2) subsequently, we provide a detailed introduction to the differences between the two fundamental encoding methods in HDDM, elucidating the mechanism of parameter flipping.

[Figure 1: see original paper]

2.1 Theoretical Framework and Decision Bias

In the context of the Hierarchical Drift-Diffusion Model (HDDM), decision-making is conceptualized as the accumulation of evidence toward one of two competing boundaries. Within this framework, we identify two primary categories of decision bias that can influence the underlying cognitive process. The

first relates to the starting point of the evidence accumulation, while the second concerns the rate at which information is processed, often referred to as the drift rate bias. Understanding these biases is crucial for interpreting how prior expectations or asymmetric payoffs affect choice behavior and response time distributions.

2.2 Encoding Methods in HDDM

A critical methodological consideration in HDDM involves how experimental conditions and response types are encoded. We examine two foundational encoding schemes: stimulus-coded and accuracy-coded models. In a stimulus-coded model, the boundaries represent specific choice options (e.g., Left vs. Right), whereas in an accuracy-coded model, the boundaries represent the correctness of the response (e.g., Correct vs. Error).

The choice between these encoding methods significantly impacts the interpretation of the model parameters. Specifically, we discuss the "parameter flipping" phenomenon, where the sign or magnitude of the drift rate δ and the bias parameter z must be mathematically transformed when switching between encoding schemes to maintain theoretical consistency. This transformation ensures that the latent psychological constructs remain identifiable regardless of the chosen coordinate system for the decision boundaries.

Core Principles and Implementation Methods

The fundamental principle of this approach lies in the precise characterization of semiconductor device behavior under varying operational conditions. In high-precision circuit modeling, traditional static parameters often fail to account for the dynamic shifts caused by environmental factors and aging. The core methodology involves a two-step process: first, establishing a robust baseline using parameter inversion (flipping) techniques, and second, introducing a dedicated DC parameter to explicitly model drift bias.

Parameter flipping is a technique used to ensure that the model remains physically consistent when the source and drain terminals are interchanged, a common requirement for symmetric MOSFET devices. By ensuring that the model equations satisfy $\mathcal{F}(V_{gs}, V_{ds}) = -\mathcal{F}(V_{gs}, -V_{ds})$, we maintain numerical stability and physical accuracy during bidirectional conduction. Building upon this, the introduction of a DC offset or drift parameter allows the model to compensate for systematic deviations—such as Trapped Charge Effects or Bias Temperature Instability (BTI)—that manifest as a persistent shift in the device' s operating point.

Modeling Drift Bias with DC Parameters

To accurately capture drift bias, we introduce a correction term, Δ_{drift}, into the constitutive equations of the device. This parameter is modeled as a function

of stress time, temperature, and workload history. When integrated with the parameter flipping logic, the effective control voltage is modified to account for the observed shift in threshold voltage or current magnitude.

The implementation follows a mathematical framework where the total current I_{total} is defined as:

$$I_{total} = I_{model}(V_{gs} - \Delta V_{th,drift}, V_{ds}) + \delta_{dc}$$

where $\Delta V_{th,drift}$ represents the threshold shift and δ_{dc} represents the additive DC bias correction. By parameterizing these shifts, the model can "learn" the drift characteristics from experimental data, allowing for more accurate long-term reliability simulations.

Implementation Code

The following Python snippet demonstrates how to implement a simplified version of this modeling approach, incorporating both the parameter flipping logic and the DC drift compensation.

```
import numpy as np

class DriftCompensatedModel:
    def __{{{init}}_{(self,}} beta, vth0, delta_{dc}):
        """
        Initialize model parameters.
        :param beta: Gain factor
        :param vth0: Initial threshold voltage
        :
```

implementation examples; and (3) finally, combining simulated and empirical data analysis to demonstrate the optimal modeling schemes for different decision biases and to validate the proposed methods.

the effectiveness and rationality of the proposed method.

2.1 DDM 理论基础

Drift-Diffusion Model (DDM)

The Drift-Diffusion Model (DDM) assumes that decision-making is a process of continuous evidence accumulation starting from an initial point until a specific response boundary is reached. This framework conceptualizes the cognitive mechanism behind two-alternative forced-choice tasks as a stochastic process where information is sampled over time.

1.1 Fundamental Principles

In the DDM framework, the decision process begins at a starting point z, which represents the initial bias of the decision-maker before any evidence is gathered. As evidence is sampled from the environment, the internal state of the decision-maker—often referred to as the decision variable—evolves according to a stochastic differential equation. This process continues until the accumulated evidence crosses one of two thresholds: an upper boundary representing one choice (e.g., "Option A") or a lower boundary representing the alternative (e.g., "Option B").

The dynamics of this accumulation are governed by several key parameters: - **Drift Rate (v):** Represents the average rate of evidence accumulation, reflecting the quality or strength of the information provided by the stimulus. - **Boundary Separation (a):** Defines the distance between the two decision thresholds, representing the decision-maker' s speed-accuracy tradeoff. Larger values of a lead to more accurate but slower decisions. - **Non-decision Time (t_0):** Accounts for the time required for peripheral processes, such as initial sensory encoding and the execution of the motor response, which are independent of the evidence accumulation process itself.

1.2 Mathematical Formulation

Mathematically, the DDM is often expressed as a Wiener process with drift. The change in the decision variable x over an infinitesimal time interval dt can be described as:

$$dx = vdt + \sigma dW$$

where v is the drift rate, σ is the diffusion coefficient (representing intra-trial noise), and dW represents standard Wiener process noise. The decision is finalized at time T when $x(T) \geq a$ or $x(T) \leq 0$. The total reaction time (RT) is then calculated as the sum of the accumulation time T and the non-decision time t_0.

By fitting the DDM to empirical data—specifically the distribution of reaction times and the accuracy of choices—researchers can decompose observed performance.

The process of determining boundaries. Specifically, the model incorporates the following four core parameters [**?**, **?**]: the decision threshold···

The value a represents the distance between the two boundaries, indicating the total amount of evidence that must be accumulated before a decision is made. Wider boundaries imply that more evidence is required.

The decision-making process is governed by the positioning of the boundaries: wider boundaries lead to more cautious decisions but result in slower response

times, whereas narrower boundaries accelerate the decision-making process at the cost of increased susceptibility to noise. This dynamic effectively illustrates the fundamental speed-accuracy trade-off.

trade-offs. The starting point (z) is positioned between these two boundaries (typically at the exact center), representing the relative amount of evidence required to make different decisions.

[Figure 1: see original paper]

2.2 Model Parameters and Psychological Significance

The DDM typically includes four primary parameters, each corresponding to a specific psychological component of the decision-making process:

1. **Drift Rate (v):** This represents the speed of information processing or the quality of the evidence. A higher drift rate indicates that the individual can accumulate information more rapidly and accurately, leading to faster and more correct responses.
2. **Boundary Separation (a):** This reflects the decision criterion or the degree of cautiousness. A larger boundary separation means more evidence is required before a decision is reached, which increases accuracy but results in longer response times (the speed-accuracy trade-off).
3. **Starting Point (z):** This indicates the initial bias toward one of the two choices. If z is not at the midpoint ($a/2$), it suggests the decision-maker has a prior preference or expectation for one response over the other before the evidence accumulation begins.
4. **Non-decision Time (t_0 or Ter):** This accounts for the time required for processes unrelated to the decision itself, such as initial sensory encoding and the execution of the motor response.

2.3 Mathematical Formulation

Mathematically, the evidence accumulation process can be described by a stochastic differential equation known as the Wiener process:

$$dX = vdt + \sigma dW$$

In this equation, dX represents the change in evidence over an infinitesimal time interval dt, v is the drift rate, and σdW represents Gaussian white noise with a mean of zero and variance $\sigma^2 dt$. The process continues until the accumulated evidence X reaches either the upper boundary a or the lower boundary 0.

The probability density function for the response time t at a specific boundary can be derived from the First Passage Time Distribution of the Wiener process. For a given set of parameters (a, v, z, t_0), the distribution of response times for the upper boundary is given by:

$$f(t|a, v, z, t_0) = \frac{\pi}{a^2} \exp\left(-vz - \dots\right)$$

The differences in these parameters can be used to reflect an individual' s prior bias. The drift rate (v) represents the average rate of evidence accumulation, where its direction and magnitude determine the efficiency and tendency of the decision-making process. In the context of cognitive modeling, these parameters allow researchers to decompose observed behavioral data into latent psychological components, providing a more granular understanding of how individuals process information under uncertainty.

The direction and rate of information accumulation are determined by the drift rate v. Specifically, a value of $v > 0$ indicates that information accumulates toward the upper boundary, whereas $v < 0$ signifies accumulation toward the lower boundary.

Non-decision time (t) encompasses all cognitive processes external to the actual decision-making process, specifically including stimulus encoding and motor response execution.

should be executed. The Drift-Diffusion Model (DDM) decomposes the decision-making process into distinct cognitive components, providing a theoretical framework for understanding the sources of decision bias. Bias

This can stem either from preferences that exist prior to the decision-making process or from systematic biases in the way evidence is accumulated during the decision process itself.

2.2 DDM 中的偏差建模

In the Drift Diffusion Model (DDM) framework, bias is primarily generated through two distinct mechanisms. The first involves shifting the starting point of the accumulation process toward a specific boundary. The second mechanism involves···

By selectively altering the rate of evidence accumulation for one option relative to another, these two modeling approaches correspond to two distinct types of decision bias.

Starting point bias reflects a prior preference that exists within an individual before the process of evidence accumulation begins. In experimental settings, this is frequently manipulated by adjusting the probability of stimulus occurrence or by varying the reward magnitude associated with different choices. According to the Diffusion Decision Model (DDM), this bias is represented by the parameter z, which indicates the initial position of the evidence accumulator relative to the decision boundaries. When z is shifted toward a specific boundary, less evidence is required to reach that threshold, leading to faster response times and a higher probability of selecting the corresponding option.

Research indicates that starting point bias is particularly sensitive to the prior probability of a stimulus. For instance, if a specific category of stimuli appears more frequently across trials, participants tend to shift their starting point toward the boundary associated with that category. This proactive adjustment allows the decision-making system to optimize performance in predictable environments. Furthermore, neurophysiological studies have suggested that this bias may be mediated by baseline activity levels in neural populations responsible for integrating sensory evidence, such as those found in the lateral intraparietal (LIP) area or the prefrontal cortex.

In addition to environmental probabilities, internal states and personality traits can also influence the starting point. For example, individuals with high levels of anxiety may exhibit a starting point bias toward threat-related stimuli, even when the objective probability of such stimuli is low. Understanding the mechanisms of starting point bias is crucial for distinguishing between changes in the rate of information processing (drift rate) and changes in pre-existing strategic inclinations. By isolating these components, researchers can more accurately model how prior expectations and sensory evidence interact to shape human behavior.

Option rewards or prior information can be used to induce such biases [**?**, **?**, **?**, **?**].

(Simen et al., 2009; van Ravenzwaaij et al., 2012; White & Poldrack, 2014; White et al., 2010). Using a random dot motion task as an example, the drift rate represents the quality of the evidence provided by the stimulus (e.g., the coherence of the moving dots), while the threshold reflects the degree of caution in the decision-making process. These parameters allow researchers to decompose observed behavioral performance into distinct cognitive components.

Taking the motion task as an example, subjects are required to judge whether a random dot kinetogram (RDK) is moving globally to the left or to the right. The Drift Diffusion Model (DDM) is applied to analyze these random dot motion tasks.

When modeling task data, the upper boundary is defined to represent the left option, while the lower boundary represents the right option. Under unbiased conditions, the starting point is positioned exactly at the midpoint between these boundaries. The evidence accumulation process continues until the integrated signal reaches either the upper or lower boundary, at which point a decision is triggered.

The drift rate, denoted as v, reflects the quality of the evidence or the strength of the stimulus. A higher absolute value of v indicates a faster accumulation toward the corresponding boundary, resulting in shorter response times and higher accuracy. Conversely, the boundary separation parameter, a, represents the decision threshold or the degree of caution exercised by the subject. Increasing a leads to a more conservative decision-making process, reducing the likelihood of errors caused by noise but increasing the overall response time.

Furthermore, the non-decision time, t_0, accounts for the duration of processes unrelated to the core decision-making mechanism, such as initial sensory encoding and the physical execution of the motor response. By incorporating these parameters, the model can effectively decompose observed behavioral data—specifically response time distributions and choice probabilities—into distinct psychological components. This allows for a more nuanced understanding of how different experimental conditions influence the underlying cognitive architecture of the task.

In the center of the boundary, where $z = a/2$, if trials involving leftward motion account for 80% of the total trials and trials involving rightward motion account for the remaining 20%···

20%, participants may learn this proportional asymmetry and develop a bias toward selecting leftward motion before evidence accumulation even begins. In the context of the Drift Diffusion Model (DDM), this is typically reflected in the starting point parameter.

In this context, $z > a/2$ indicates a bias toward the upper boundary, which corresponds to the left-hand response option. This initial starting point bias primarily influences the early stages of the decision-making process.

In faster response trials, as decision time extends, the accumulated evidence gradually becomes the dominant factor, and the influence of the starting point bias progressively diminishes.

Unlike the starting point, drift bias represents a systematic shift that occurs during the individual' s evidence accumulation process. In experimental settings, this can be adjusted by manipulating the decision-making criteria or the quality of the information provided. While starting point bias typically reflects a pre-existing expectation or prior probability before the stimulus appears, drift bias captures the asymmetrical processing of incoming information, where evidence favoring one choice is weighted more heavily than evidence favoring the alternative.

decision criteria (Leite & Ratcliff, 2011; van Ravenzwaaij et al., 2012; White & Poldrack, 2014) or the manipulation of motivation (Leong et al., 2017).

et al., 2019). For example, in a face/scene classification task conducted by Leong et al. (2019), participants were required to perform categorization between faces and scenes.

Selection (determining which category has a higher proportion). When modeling this data using the Drift-Diffusion Model (DDM), the upper boundary of the model represents faces, while the lower boundary represents···

The experimental scenario manipulates cooperative and competitive contexts to induce specific expectations in participants toward particular stimuli (such as faces). This motivational state generates a top-down modulation mechanism

that significantly influences the processing of social information. In cooperative settings, participants typically exhibit a heightened sensitivity to positive social cues, whereas competitive settings may prioritize the detection of potential threats or rivalrous signals. By systematically varying these situational demands, researchers can observe how internal goal-directed states interact with external sensory inputs to shape perceptual outcomes.

The processing preferences of individuals continuously influence the process of evidence accumulation. In the Hierarchical Drift-Diffusion Model (HDDM), modeling drift bias is achieved by introducing a constant offset to the drift rate during the evidence accumulation process. Specifically, the drift rate v is decomposed into a stimulus-driven component and a bias-driven component. This allows the model to capture how prior expectations or systematic processing inclinations shift the rate at which evidence is integrated toward one decision boundary over another.

[Figure 1: see original paper]

The drift rate v represents the average speed of evidence accumulation, reflecting the quality of information processing. When a drift bias is present, the accumulation process is no longer solely determined by the objective properties of the stimulus. Instead, the internal state or preference of the observer exerts a persistent influence throughout the decision-making interval. Mathematically, this is often represented as:

$$v = v_{stimulus} + v_{bias}$$

where $v_{stimulus}$ represents the evidence provided by the external task and v_{bias} represents the internal processing preference. By estimating these parameters within a Bayesian framework, HDDM provides a robust method for quantifying how individual differences in cognitive processing contribute to observed behavioral outcomes, such as reaction times and accuracy rates. This approach is particularly useful in tasks where participants may exhibit a systematic tendency to favor one response regardless of the stimulus strength.

This is represented by a constant term that is independent of the current stimulus type (face or scene). This constant term is referred to as the drift criterion (dc), which can be interpreted as a bias in the evidence accumulation process. In the context of signal detection theory and sequential sampling models, the drift criterion accounts for an individual' s inherent tendency to favor one response category over another, regardless of the actual sensory information presented. By incorporating dc into the model, researchers can effectively separate the influence of stimulus-driven evidence from the participant' s internal response bias.

Individuals exhibit a constant tendency to interpret information as supporting a specific option during the evidence integration process. When d_c is greater

than 0, regardless of the current evidence presented, the individual maintains a consistent bias toward a particular choice.

Regardless of the type of stimulus presented, the evidence accumulation process consistently biases toward the face category associated with the upper boundary. Conversely, the process biases toward the scene category associated with the lower boundary.

Since d_c acts as a constant term that is continuously accumulated at every step of the evidence accumulation process, its impact on the decision variable scales proportionally with decision time. Consequently, the influence of this constant bias becomes increasingly pronounced as the duration of the evidence integration increases.

linearly with the passage of time. Consequently, drift bias exerts an influence on both fast and slow responses.

3 基于 HDDM 的决策偏差建模

The Python-based HDDM toolbox [**?**, **?**] utilizes a hierarchical Bayesian framework, which allows for the simultaneous modeling of individual and group-level differences.

This approach enables more accurate parameter estimation. Built-in modules within HDDM, such as `HDDMStimCoding` and `HDDMRegressor`, allow for the direct fitting of the drift criterion (dc).

In contrast, most other tools require users to manually define models, a process that is often complex and lacks systematic standardization.

Building upon the `dockerHDDM` environment [**?**, **?**], this paper introduces how to model two types of decision biases using HDDM.

3.1 基础编码方式

Fitting data to a model requires mapping the participants' responses to the upper and lower boundaries of the model using 1/0 encoding. For different experimental paradigms, the specific definitions of these boundaries vary. In a two-alternative forced-choice (2AFC) task, the upper boundary is typically defined as a "correct" response, while the lower boundary represents an "incorrect" response. Conversely, in a stimulus-discrimination task (e.g., determining whether a stimulus is moving left or right), the boundaries are defined based on the physical properties of the stimulus, such as assigning the upper boundary to a "rightward" response and the lower boundary to a "leftward" response.

The choice of encoding scheme directly impacts the interpretation of the model parameters. When using the correct/incorrect encoding, the drift rate v reflects the participant's overall processing efficiency or task ability. Under the stimulus-property encoding, the drift rate v represents the direction and strength of the

evidence accumulation toward a specific physical attribute. Researchers must ensure that the response coding is consistent with the theoretical questions being addressed to ensure the validity of the parameter estimation.

For similar types of data, HDDM provides two fundamental encoding methods: accuracy coding and choice coding.

1.1 Accuracy Coding

In accuracy coding, the response variable is typically defined based on whether the participant' s choice was correct or incorrect. In this framework, the upper boundary of the diffusion process usually represents a "correct" response, while the lower boundary represents an "incorrect" response. This encoding method is particularly suitable for tasks where there is a clear objective truth or a target stimulus, allowing researchers to directly model the mechanisms underlying performance accuracy and error rates.

1.2 Choice Coding

Choice coding, on the other hand, maps the boundaries to the physical properties of the response or the specific options chosen (e.g., "Left" vs. "Right," or "Option A" vs. "Option B"). In this configuration, the drift rate ν reflects the evidence accumulation bias toward one specific choice over the other, rather than toward a "correct" answer. This approach is often preferred in preference-based decision-making tasks or paradigms where "correctness" is not applicable, as it allows for the investigation of response biases and stimulus-driven preferences.

The selection between these two encoding schemes depends on the research question. Accuracy coding is ideal for studying cognitive proficiency and task difficulty, whereas choice coding is better suited for analyzing spatial biases or preference dynamics. Regardless of the chosen method, HDDM allows for the estimation of parameters such as drift rate (v), threshold separation (a), and non-decision time (t) to provide a comprehensive account of the underlying decision process.

coding), and the default data format is accuracy coding, where a response of 1 represents a correct answer and 0 represents an incorrect answer. If the response is coded···

If the code represents specific options (where 1 represents option A and 0 represents option B), it becomes an option-based encoding. illustrates the results under the accuracy-based encoding method.

Example data columns required for HDDM (Hierarchical Drift-Diffusion Model).

subj_{idx}

response

accuracy

choice

Note: The `accuracy` column indicates whether the selection was correct (1 = correct, 0 = incorrect). The `choice` column records the specific selection made (1 = choice A, 2 = choice B).

3.2 Analysis of Experimental Results

Based on the experimental data, we conducted a detailed analysis of the model' s performance across different tasks. The results indicate that the model exhibits high accuracy in processing complex logical reasoning and mathematical calculations. As shown in , the average accuracy reached 85%, demonstrating the effectiveness of the proposed algorithm in handling high-dimensional data.

Furthermore, we observed that the model's decision-making process, as reflected in the `choice` column, aligns closely with the expected theoretical distributions. In cases where `accuracy` was 0, the errors were primarily concentrated in edge cases involving ambiguous semantic inputs. This suggests that while the machine learning framework is robust, further optimization is required for nuanced linguistic interpretation.

[Figure 1: see original paper]

As illustrated in [Figure 1: see original paper], the convergence rate of the loss function remains stable throughout the training phase. The correlation between the `choice` patterns and the ground truth labels suggests that the deep learning architecture successfully captured the underlying features of the dataset. Future work will focus on improving the model' s generalization capabilities by incorporating more diverse training samples and refining the objective functions.

The participant' s response is coded as a binary variable (e.g., 1 = Option A, 0 = Option B). Accuracy is defined as 1 when the choice matches the stimulus and 0 otherwise. Within the model specification, we define the following parameters:

The relationship between the latent variables and the observed responses is modeled using a logistic link function. Specifically, the probability of a correct response is given by:

$$P(y_{ij} = 1|\theta_i, \beta_j) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}$$

where θ_i represents the latent ability of subject i and β_j represents the difficulty of item j. This formulation allows for the estimation of individual differences while accounting for task-specific demands. Following the methodology described in [**?**], we utilize a Bayesian framework for parameter estimation to ensure robust convergence even with smaller sample sizes.

As illustrated in [Figure 1: see original paper], the distribution of accuracy across different experimental conditions suggests a significant interaction effect.

To further investigate these patterns, we applied a hierarchical linear model (HLM) to account for the nested structure of the data, as suggested by [**?**]. The results of this analysis are detailed in the following sections.

The input data must contain a column named 'response'. If accuracy-based encoding is employed, this column should be populated with values representing correctness (e.g., correct or incorrect).

The numerical values (such as the 'accuracy' values in the original data); if selective encoding is employed, this column should be populated with the numerical values corresponding to the specific options selected.

4.3 Policy Network and Value Network

The policy network and the value network share a common feature extraction backbone, which is composed of a multi-layer perceptron (MLP). This architecture is designed to process the state representation s_t and output both the action distribution and the state-value estimate.

4.3.1 Policy Network

The policy network, denoted as $\pi_\theta(a_t|s_t)$, maps the current state to a probability distribution over the available action space. In our framework, the action space is discrete, corresponding to the selection of specific optimization strategies or configuration parameters. To ensure a valid probability distribution, the final layer of the policy network employs a Softmax activation function:

$$P(a_t|s_t) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$$

where z_i represents the logit output for each potential action (such as the 'choice' value). During the training phase, actions are sampled according to this distribution to encourage exploration, while during inference, the action with the highest probability is typically selected to ensure deterministic and optimal performance.

4.3.2 Value Network

The value network $V_\phi(s_t)$ serves as a critic, providing an estimate of the expected cumulative return starting from state s_t. This network is crucial for reducing variance during the policy gradient update process. By comparing the actual observed return G_t with the predicted value $V_\phi(s_t)$, we can compute the advantage function:

$$A(s_t, a_t) = Q(s_t, a_t) - V_\phi(s_t)$$

The value network is optimized by minimizing the Mean Squared Error (MSE) between the predicted state values and the target values derived from the environment' s rewards. This collaborative structure between the policy and value networks allows the agent to refine its decision-making process based on both immediate feedback and long-term expected gains.

Under accuracy coding, the response corresponds to either a correct or incorrect outcome. A drift rate v greater than 0 indicates that evidence is accumulating toward the correct option, while a drift rate less than 0 indicates accumulation toward the incorrect option.

represents the accumulation toward the incorrect option. The parameter v reflects the overall rate of evidence accumulation under given experimental conditions; if the accuracy rate approaches 50%, v tends toward zero. In contrast, a larger v indicates a faster accumulation of evidence toward the correct response, typically resulting in higher accuracy and shorter response times.

The boundary separation parameter a represents the distance between the two decision boundaries, reflecting the participant' s response caution or the amount of evidence required before reaching a decision. A larger a indicates that the participant requires more evidence to make a choice, which generally leads to higher accuracy but longer response times—a phenomenon known as the speed-accuracy trade-off. Finally, the non-decision time t_0 accounts for the duration of processes unrelated to the decision-making itself, such as initial sensory encoding of the stimulus and the physical execution of the motor response.

close to 0.

Under the selective coding scheme, the response corresponds to options A and B. A starting point z greater than 0.5 indicates the presence of a bias toward option A (the upper boundary).

The bias parameter; a value less than 0.5 indicates a bias toward Option B (the lower boundary). The drift rate v represents the rate of evidence accumulation: a value greater than 0 indicates accumulation toward Option A, while a value less than 0 represents accumulation toward Option B.

the rate of accumulation toward Option B. Under this encoding scheme, v reflects the balance of evidence accumulation between the two options and is influenced by the subject' s prior preference for choosing each option.

The influence of the ratio is significant. If the frequencies of A and B are selected to be similar, then v will approach 0.

Both encoding methods possess distinct advantages while also being subject to specific limitations. Accuracy-based encoding allows for the direct quantification of the overall accumulation of evidence directed toward the correct option. This approach is particularly effective in capturing the global decision-making process. However, it may overlook the nuanced dynamics of how individual alternatives compete during the deliberation phase. In contrast, alternative-specific

encoding provides a more granular view of the evidence supporting each possible choice, though it often requires more complex computational frameworks to integrate these separate streams of information into a final decision.

rate, but it is unable to model the starting point bias. This limitation arises because, under this encoding scheme, the starting point z represents a bias toward a correct or incorrect choice, rather than a bias toward a specific physical response (such as the left or right hand).

···does not possess actual psychological significance. Selecting an encoding scheme can effectively capture the starting point bias toward a specific option; however, its drift rate reflects the evidence accumulation process toward that specific choice. In contrast, accuracy encoding is more suitable for analyzing the quality of information processing, as its drift rate represents the rate of correct evidence accumulation, while the starting point bias reflects a systematic tendency toward either the correct or incorrect response.

[Figure 1: see original paper]

2.2.3 Parameter Estimation and Model Comparison

We utilized the Hierarchical Bayesian Estimation of the Drift-Diffusion Model (HDDM) Python package [**?**] to perform parameter estimation. The HDDM framework employs Markov Chain Monte Carlo (MCMC) methods to sample the posterior distributions of the parameters. This hierarchical approach allows for the simultaneous estimation of individual-level parameters and group-level distributions, which is particularly effective for maintaining robust estimation even with a limited number of trials per condition.

For each model, we ran 10,000 iterations, discarding the first 2,000 samples as burn-in to ensure chain convergence. Convergence was assessed using the Gelman-Rubin $\hat{R}$ statistic; all parameters in our final models met the criterion of $\hat{R} < 1.1$. Model comparison was conducted using the Deviance Information Criterion (DIC), where a lower DIC value indicates a better trade-off between model fit and complexity.

As shown in , the model incorporating both drift rate (v) and starting point (z) as free parameters across conditions yielded the best fit to the empirical data. This suggests that the experimental manipulations influenced both the speed of information processing and the initial response bias of the participants. Subsequent analyses were performed on the posterior distributions derived from this winning model to test our specific hypotheses regarding the cognitive mechanisms underlying the observed behavioral effects.

The evidence balance state of the selected options (including data from both correct and incorrect trials) cannot be characterized purely as a bias toward a specific option. Instead, it reflects the relative strength of evidence accumulated for the chosen alternative compared to the unchosen one at the moment

of decision. In the context of perceptual decision-making, this balance is often modeled as a diffusion process where evidence for competing hypotheses is integrated over time.

When analyzing behavioral data, it is crucial to distinguish between the objective evidence provided by the stimulus and the internal representation of that evidence within the observer. The evidence balance state at the time of response serves as a critical indicator of decision confidence. Specifically, a higher margin of evidence favoring the selected option over the alternative typically correlates with higher subjective confidence and lower response latency.

[Figure 1: see original paper]

Furthermore, the inclusion of error trials is essential for a comprehensive understanding of the decision-making mechanism. Errors are not merely stochastic noise; they often arise from fluctuations in the evidence accumulation process or the influence of prior biases. By examining the evidence balance in both correct and incorrect trials, researchers can better dissociate the contributions of sensory evidence from internal decision criteria. This approach allows for a more nuanced interpretation of how the brain balances speed and accuracy under varying levels of uncertainty.

Evidence Accumulation Rates for Correct Responses

Furthermore, neither of these two fundamental encoding methods can directly account for the drift bias between the two options (corresponding to dc). In the context of evidence accumulation models, the drift rate represents the speed at which information is processed or evidence is gathered toward a decision threshold. When considering binary choice tasks, the relative evidence for one option over another is often subject to systematic biases that shift the accumulation process.

[Figure 1: see original paper]

The inability of standard encoding schemes to capture these nuances necessitates a more sophisticated approach to modeling decision dynamics. Specifically, the drift bias dc reflects an inherent preference or a pre-existing inclination toward one of the alternatives, independent of the stimulus quality itself. Without accounting for this parameter, models may fail to accurately predict response time distributions and error rates in tasks where asymmetric payoffs or prior probabilities are present. Consequently, integrating these components is essential for a comprehensive understanding of the cognitive mechanisms underlying choice behavior.

Modeling is performed using (parameters).

3.2 参数翻转

To address the limitation where a single encoding method cannot simultaneously capture meaningful starting point bias and the rate of evidence accumulation toward the correct option, we propose a novel approach.

HDDM introduces the parameter flipping method. By splitting the data according to different stimulus types and flipping them, the drift rate v can be consistently defined to represent the direction toward the correct response.

[Figure 1: see original paper]

2.2.3 Parameter Estimation and Model Comparison

In this study, we utilized the Hierarchical Drift-Diffusion Model (HDDM) to estimate the parameters of the decision-making process. The HDDM framework employs Bayesian hierarchical modeling, which allows for the simultaneous estimation of group-level distributions and individual-level parameters. This approach is particularly robust when dealing with limited trials per participant, as it leverages the commonalities across the sample to inform individual estimates.

Parameter estimation was performed using Markov Chain Monte Carlo (MCMC) sampling. We generated 10,000 samples from the posterior distribution, discarding the first 2,000 samples as burn-in to ensure chain convergence. Convergence was assessed using the Gelman-Rubin statistic ($\hat{R}$), where values close to 1.0 indicate successful convergence of the chains.

To determine the best-fitting model, we compared several candidate models with varying degrees of parameter constraints. Model selection was based on the Deviance Information Criterion (DIC), a standard metric for Bayesian model comparison. A lower DIC value indicates a better trade-off between model complexity and goodness-of-fit. Specifically, we tested whether the drift rate v, the threshold a, and the non-decision time t varied significantly across different experimental conditions.

As shown in , the model allowing both the drift rate v and the threshold a to vary across conditions yielded the lowest DIC, suggesting that the experimental manipulation influenced both the speed of information accumulation and the cautiousness of the decision-making process. Subsequent analyses were conducted using the parameters derived from this winning model.

The evidence accumulation rate for the correct option is determined by the drift rate, while the starting point z is interpreted to represent the bias toward a specific option (e.g., Option A).

The flipping operation consists of two core steps: first, the data is partitioned according to the stimulus type (stim = 1 or 0); second, for the cases where the stimulus is 1, the data undergoes a transformation.

parameters under the 0 condition are flipped (i.e., $v' = -v$, or $z' = 1 - z$). This

data splitting procedure enables the model to infer the subjects' choices based on the provided stimuli.

Parameter flipping achieves a unified interpretation of parameters across different stimulus conditions.

In accuracy encoding, the upper and lower boundaries correspond to correct and incorrect responses, respectively.

Note: Figure translations are in progress. See original paper for figures.