Response Time Enhances Alignment with Heterogeneous Preferences

Federico Echenique
Department of Economics
University of California, Berkeley

Alireza Fallah
Department of Computer Science
and Ken Kennedy Institute
Rice University

Baihe Huang
Departments of Electrical Engineering
and Computer Sciences
University of California, Berkeley

Michael I. Jordan
Departments of Electrical Engineering
and Computer Sciences and Statistics
University of California, Berkeley;
Inria Paris

Abstract

Aligning large language models (LLMs) to human preferences typically relies on aggregating pooled feedback into a single reward model. However, this standard approach assumes that all labelers share the same underlying preferences, ignoring the fact that real-world labelers are highly heterogeneous and usually anonymous. Consequently, relying solely on binary choice data fundamentally distorts the learned policy, making the true population-average preference unidentifiable. To overcome this critical limitation, we demonstrate that augmenting preference datasets with a simple, secondary signal—the user's response time—can restore the identifiability of the population's average preference. By modeling each decision as a Drift-Diffusion Model (DDM), we introduce a novel, consistent estimator of heterogeneous preferences that successfully corrects the distortions of standard choice-only labels. We prove that our estimator asymptotically converges to the true average preference even in extreme cases where each anonymous labeler contributes only a single choice. Empirically, across both synthetic and real-world datasets, our method consistently outperforms standard baselines that otherwise fail and plateau at a bias floor. Because response times are essentially free to record and require zero user tracking or identification, our results bring promises and open up new opportunities for future data-collection pipelines to improve the social benefit without requiring user-level identifiers or repeated elicitations.

1 Introduction

Large language models (LLMs) are aligned to human preferences through pairwise-comparison data [46, 42, 26, 4]. The standard methodology, reinforcement learning from human feedback (RLHF) [15, 67, 47, 6], fits a Bradley–Terry reward model to binary preferences and then optimizes the policy against that reward. Direct Preference Optimization (DPO) [51] and related objectives [5, 21, 43, 31] change the optimization layer, but they keep the same single-reward statistical backbone.

These methods typically assume a unique preference shared by all labelers. Real-world datasets, however, pool anonymous annotators with highly heterogeneous tastes, values, and expertise. Recent work shows that this mismatch can badly distort the learned policy and can even make the population-average preference unidentifiable from choices alone [65, 64, 28, 12, 49].

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

One solution is to estimate each labeler's preferences separately, but this presents two significant challenges. First, large-scale datasets are often anonymous or contain too few comparisons per labeler. Second, even with perfect access to each labeler's individual preferences, how are we to aggregate them into a single aggregate reward?

We deal with the first problem by jointly modeling a group of labelers. Each individual is assumed to follow the *Drift Diffusion Model* (DDM), the dominant procedural choice model in psychology and neuroscience [52, 54, 27, 22, 16]. The DDM assumes a Brownian motion with latent drift V absorbed at $\pm b$, where V is the signed utility difference between the two choices, and b is the decision threshold. Different labelers have different utilities and thus idiosyncratic drifts, but they share the boundary parameter b.

We deal with the second problem using response time to gauge utility intensity. The problem of uncovering a common aggregate reward has a long history in social choice theory. The 18th-century philosopher Marquis de Condorcet famously noted that pairwise comparisons cannot always be aggregated into a single winning choice. The modern version of Condorcet's paradox is Arrow's impossibility theorem, which states that no voting procedure satisfies basic efficiency and fairness properties.

Our solution to the paradoxes of Condorcet and Arrow lies in the measurement of utility. d'Aspremont and Gevers [18] show that Arrow's theorem stems from the assumption that utilities are entirely ordinal, making them incomparable. By introducing degrees of comparability, they demonstrate that a utilitarian criterion (such as an average reward) can emerge. This comparability, however, requires a mechanism to measure the intensity of preferences. Our approach rests precisely on such a mechanism.

The DDM predicts that response time and utility are related. An increase in the intensity of utility translates into a larger drift, and therefore a shorter response time. Indeed, marginalizing out response time reduces the DDM directly to the Bradley–Terry model. Recent work shows that, when b is known, response times recover preference parameters at the parametric $1/n$ rate [19]. Our departure is the practically relevant case in which the boundary is *common but unknown*. We therefore estimate the latent boundary and the population-average drift jointly from anonymous choices and response times. Our contributions are three-fold.

- **Unbiased drift and consistent boundary estimation.** Given the boundary b, we derive an unbiased estimator of the utilitarian aggregate reward, defined as the average drift $\mu = \mathbb{E}[V]$. The estimator is a weighted average of response times, with closed-form weights $w_b(\cdot)$. From the large-λ tail of the response-time Laplace transform, we then construct a Richardson-extrapolated consistent estimator $\widetilde{B}_n$ of b, whose population bias is $O(\lambda_n^{-1})$—improving the one-scale $O(\lambda_n^{-1/2})$ bias under bounded drifts.
- **Consistent average preference in the linear model.** In the heterogeneous linear preference model with per-labeler coefficient $\theta_i \in \mathbb{R}^d$ and drift $V_i = \psi_i^\top \theta_i$, combining the plug-in response-time weight $w_{\widetilde{B}_n}(\cdot)$ with ordinary least squares yields a consistent estimator of the population-average preference vector $\theta^\star = \mathbb{E}[\theta_i]$, with asymptotically negligible plug-in error from $\widetilde{B}_n$.
- **Empirical validation.** On synthetic data over diverse latent-preference priors and on an intertemporal-choice dataset from Amasino et al. [3], reanalyzed by Echenique et al. [19], the response-time plug-in matches an oracle that knows the boundary, while the choice-only baseline plateaus at a heterogeneity-induced bias floor.

To sum up, the use of response time in an anonymous, heterogeneous population of decision makers allows for the recovery of a population-average preference that serves as a utilitarian aggregate objective. While response times are not yet logged in standard preference-learning datasets for LLM alignment, the required instrumentation is minimal; our results strongly argue for recording this essential free signal in future data-collection pipelines.

1.1 Related Work

Reinforcement learning from human feedback. RLHF fits a reward model to binary preferences via the Bradley–Terry likelihood and then optimizes the policy against that reward [15, 67, 47, 6, 77, 8, 61]. DPO and related objectives replace the optimization layer while keeping the same single-reward statistical target [51, 5, 74, 21, 43, 31, 70]. A complementary line of work studies reward-model

over-optimization, robustness, and the gap between online and offline alignment [25, 76, 75, 20, 14, 41, 69, 11]. Response times do not enter the statistical design of these methods.

Heterogeneous preferences and the Bradley–Terry distortion. Recent work shows that a single Bradley–Terry reward can be badly misspecified under heterogeneous feedback. Hidden-context analyses imply Borda-style aggregation rather than mean aggregation [65]; direct alignment without per-labeler information faces a consistency-sample-efficiency tradeoff [64]; and social-choice analyses quantify severe worst-case distortion for Bradley–Terry-based alignment, with Nash Learning from Human Feedback emerging as distortion-optimal in the worst case [28, 45]. Other approaches enrich the learner with mixtures, personalization, or explicit value axes [12, 49, 66]. We instead keep the learner simple and add response time as an orthogonal signal, with no labeler identifier at all.

Social-choice criteria and probabilistic voting. Xiao et al. [72] characterize preference profiles under which reward-based alignment satisfies pairwise-majority and Condorcet consistency. Liu et al. [40] study the statistical obstacles posed by Condorcet cycles and the conditions under which Nash-based alignment preserves mixed strategies. Shi et al. [63] further characterize Condorcet and Smith consistency for game-theoretic alignment and establish limits on preference matching. Distortion under probabilistic voting is studied in the metric setting by Goyal and Sarmasarkar [29] and in the utilitarian setting by Alipour and Goyal [1]. In particular, the latter bound the utilitarian distortion of Copeland and Borda by $\beta(1+e^{-\beta})/(1-e^{-\beta})$, independently of the number of alternatives, where β is the inverse temperature in their Plackett–Luce model. These results characterize social-choice guarantees from ordinal feedback; our response-time observations identify the cardinal population-average utility under the DDM.

Drift-diffusion model and response-time-based inference. The DDM is a standard model of perceptual and value-based binary choice [52, 55, 54, 27, 35, 22, 68]. It yields explicit formulas for the joint law of choice and first-passage time and has been used in food choice, intertemporal choice, advertising, and laboratory decision tasks [35, 3, 13, 16, 23, 24]. On the inference side, prior work develops nonparametric tests, ordinal inference from response times, and $1/n$-rate estimation when the boundary is known [24, 2, 19]; recent ML work also studies response times in robust preference learning and linear bandits [60, 38]. To our knowledge, our paper is the first response-time-based preference-learning work that simultaneously targets the population-average preference in an anonymous heterogeneous labeler pool and provides a consistent estimator when the common DDM boundary is unknown and must be estimated from the same response-time data.

Response-time variation and online data. The treatment of response-time outliers and contaminant latencies has a substantial methodological literature [53, 56]. Online applications of evidence-accumulation models include goal-directed decisions [39], social information and charity preferences [36, 58], and studies of the speed–accuracy tradeoff and the interpretation of drift rates [30, 71]. Copeland et al. [17] and Ozkan and Zhang [48] study how task wording and information sources affect value-based evidence accumulation using online experiments. These applications motivate distinguishing evidence variability from non-decision latency, as in Remark 1 and Appendix B.1. For intertemporal choice, Pongratz and Schoemann [50] compile a database of choices and response times across 100 studies; Appendix C.7 evaluates our estimator on six eligible studies from this collection.

2 Preliminaries

Preference learning. A preference dataset is a set of tuples $\mathcal{D} = \{(x_i, y_i, z_i)\}_{i=1}^n$ in which $(x_i, y_i) \in \mathcal{X}^2$ is a pair of alternatives (e.g. two LLM responses to the same prompt) and $z_i \in \{+1, -1\}$ encodes whether labeler i prefers x_i ($z_i = +1$) or y_i ($z_i = -1$). The standard Bradley–Terry model assumes a latent reward $r_\theta : \mathcal{X} \to \mathbb{R}$ and posits

$$\mathbb{P}(z_i = +1 \mid x_i, y_i) = \sigma\big(r_\theta(x_i) - r_\theta(y_i)\big), \qquad \sigma(u) = \tfrac{1}{1+e^{-u}}.$$

Classical RLHF [47] first fits θ by maximum likelihood and then optimizes a KL-regularized objective against r_θ. Variants such as DPO [51], IPO [5], KTO [21], SimPO [43], and ORPO [31] retain the single-reward assumption but differ in the loss or the parameterization.

Alignment under heterogeneous preferences. Real-world preference datasets pool annotations from many labelers with potentially different preferences. Label i comes from an individual labeler with utility $u_i : \mathcal{X} \to \mathbb{R}$, drawn i.i.d. from a population distribution. Individual i is never observed

again. We seek to learn the *average utility,* which corresponds to the utilitarian criterion we discussed in the introduction. Siththaranjan et al. [65], Shirali et al. [64], Gölz et al. [28] show that the single-reward likelihood is misspecified under heterogeneity: even with infinitely many samples, the Bradley–Terry MLE converges to the *Borda-rule* aggregate rather than the average-utility. The gap between the utilitarian and Borda aggregates can be large (see Section 4.1 for details), and a utilitarian goal has stronger foundations when utility is cardinally meaningful. Consequently, a model trained with RLHF or DPO on heterogeneous data need not rank alternatives in the order dictated by the utilitarian criterion $\mathrm{AvgUtil}(x) := \mathbb{E}[u_i(x)]$. Shirali et al. [64] prove an incompatibility result between consistency and sample efficiency when per-labeler information is unavailable, motivating the need to go beyond binary choice data: hence response times.

Drift-diffusion model. The Drift diffusion model (DDM) [52, 54] is the dominant model in psychology and neuroscience for modeling choice and response time. The DDM assumes that information about x_i and y_i arrives over time according to a Brownian motion W_τ with drift v. When W_τ first reaches a boundary b or $-b$ a decision is made:

$$W_\tau = B_\tau + v \cdot \tau, \quad T = \inf\{\tau > 0 : |W_\tau| = b\}, \quad Z = 2 \cdot \mathbf{1}(W_T = b) - 1 \in \{+1, -1\}, \quad (1)$$

where (B_τ) is a standard Brownian motion with $B_0 = 0$. The sign Z encodes the choice between x_i and y_i, the stopping time T is the response time[1], and v is the underlying utility difference between x_i and y_i. The DDM can be viewed as a decision maker learning about the relative quality of two alternatives through separate Gaussian signals. Posterior beliefs become more accurate over time, but the decision maker must choose when to stop, since waiting entails a constant marginal cost [23].

Importantly, the DDM recovers the Bradley–Terry as a special case when response time is ignored. The probability of $Z = z$ in the DDM is given by $\sigma(2bzv)$ [19]. In other words, the DDM provides a joint model of choice and response time that is consistent with the Bradley–Terry and has strong support in the psychology literature [54, 57]. Furthermore, under fixed (v, b), it is known that Z and T are conditionally independent given (v, b), and $\mathbb{E}[Z \mid v, b] = \tanh(bv)$ [19, 24].

Remark 1 (Sources of response-time variation)**.** *Variation during evidence accumulation and latency outside the decision process enter the model through different components. Moment-to-moment fluctuations in evidence are represented by the Brownian term in Eq.* (1)*, while trial-to-trial variation in the quality of evidence can be represented by heterogeneous drifts. An interruption that pauses the task while preserving the eventual comparison contributes an additive non-decision time, so that the observed time is $S = T + A$. Contaminant response times and parameter variability are established subjects of response-time analysis [53, 56]. We develop the correction for a common additive delay in Appendix B.1 and independent uniform delays in Remark 5.*

Roadmap. We assume that individual labelers follow the DDM model, each with their idiosyncratic utility and, thus, an idiosyncratic drift. We seek to estimate the mean drift, which corresponds to the utilitarian criterion. We assume that all labelers share a common boundary b, which is tantamount to assuming a common temperature in Bradley–Terry. In Appendix B.2, we discuss what happens if this assumption does not hold and propose possible adjustments.

In Section 3, we start off with a simple scenario: every observation comes from a fresh labeler with an unknown latent drift $V \sim F^*$ (i.i.d.). In other words, for a labeler i with drift v_i, we only get to see (z_i, t_i) generated by Eq. (1). We just assume bounded support, meaning there is a known constant $M < \infty$ such that $|V| \leq M$ almost surely[2]. One can interpret this as asking all labelers to choose between the same two alternatives, x and y. Here, v_i is simply $u_i(x) - u_i(y)$, where $u_i(\cdot)$ represents labeler i's utility. Our main goal here is to learn the average utility difference, i.e., $\mu := \mathbb{E}_{V \sim F^*}[V]$.

In Section 4, we extend this analysis to a contextual linear setting with multiple alternatives. More precisely, labeler i is offered two alternatives $(x_i, y_i) \in \mathcal{X}^2$. We assume that labeler i's utility is linear and given by $u_i(x) = \phi(x)^\top \theta_i$, where $\theta_i \in \mathbb{R}^d$ is the labeler's unknown latent preference vector, drawn from a distribution G^*, and $\phi : \mathcal{X} \to \mathbb{R}^d$ is the known feature map. In this case, the drift becomes $v_i = (\phi(x_i) - \phi(y_i))^\top \theta_i$, and our goal is to learn the population-average preference vector $\theta^\star := \mathbb{E}_{\theta \sim G^*}[\theta]$. Note that we never observe the labeler's identity.

[1] An extended version of the DDM allows the time it takes humans to process the alternatives, called *non-decision time*, to be included in the response time [56]. In Appendix B.1, we discuss how our results can be adjusted to accommodate this.

[2] We relax this assumption to a bounded exponential-moment condition in Appendix B.3, which holds for Gaussian drift.

3 Estimating the Common Boundary and the Average Drift

Our goal is to estimate the utilitarian objective, the average drift $\mu = \mathbb{E}_{V\sim F^*}[V]$, assuming access to the dataset $(z_i, t_i)_{i=1}^n$, where z_i and t_i denote the choice and response-time data, respectively, generated by a DDM with drift $v_i \sim F^*$. We present a two-stage construction. First, we design an unbiased estimator of μ, *assuming that the boundary parameter b is known*. Next, we develop a *consistent* Richardson-extrapolated estimator of the common boundary.

3.1 Unbiased estimation of the average drift given the boundary

For $k \geq 0$ and $b > 0$, write $c_k(b) := (2k+1)b$ and for $t > 0$ define the ratio-of-series weight

$$w_b(t) \;:=\; \frac{\displaystyle\sum_{k=0}^{\infty}\left(\frac{c_k(b)^2}{t} - 1\right)\exp\left(-\frac{c_k(b)^2}{2t}\right)}{\displaystyle\sum_{k=0}^{\infty}(-1)^k c_k(b)\exp\left(-\frac{c_k(b)^2}{2t}\right)}. \tag{2}$$

Theorem 1 (Unbiased estimator of the average drift given the boundary)**.** *Fix the value of the boundary $b > 0$. Under the DDM data-generating process of Section 2, the estimator*

$$\widehat{\mu}_n \;=\; \frac{1}{n}\sum_{i=1}^{n} z_i\, w_b(t_i)$$

satisfies $\mathbb{E}[\widehat{\mu}_n] = \mu$ for every drift distribution F^*. Moreover, w_b is the unique *smooth function satisfying $\mathbb{E}[Z w_b(T) \mid V = v] = v$ for all v.*

Proof sketch: First, due to the independence between the choice Z and response time T, conditional on the drift $V = v$, $\mathbb{E}[Z w_b(T) \mid V = v]$ is equal to $\mathbb{E}[Z \mid V = v] \cdot \mathbb{E}[w_b(T) \mid V = v]$ where the first expectation is known as $\tanh(bv)$. Then, $\mathbb{E}[w_b(T) \mid V = v] = \frac{v}{\tanh(bv)}$ follows from Girsanov's theorem and Laplace transform of $\frac{\sqrt{2s}}{\sinh(b\sqrt{2s})}$. A proof is provided in Appendix A.1. □

3.2 Consistent estimation of the common unknown boundary

The weight function (2) requires knowledge of the boundary parameter b: We now turn to a consistent estimator of b. To this end, we use the empirical Laplace transform of the response times, given by $\widehat{L}_n(\lambda) := \frac{1}{n}\sum_{i=1}^n e^{-\lambda t_i}$. To see why this is useful, let us consider for a moment the homogeneous case with fixed drift $V = v$. Note that, as $n \to \infty$, the empirical Laplace transform converges to its population counterpart, $L_b(\lambda) := \mathbb{E}[e^{-\lambda T}]$, which is given by [see 19, Lemma 4]:

$$L_b(\lambda) = \frac{\cosh(bv)}{\cosh\left(b\sqrt{2\lambda + v^2}\right)}. \tag{3}$$

Consequently, as $\lambda \to \infty$, we see that $\frac{-\log L_b(\lambda)}{\sqrt{2\lambda}}$ converges to b, allowing us to recover the boundary parameter. The next theorem formalizes this idea while accounting for raters' heterogeneity in v.

Theorem 2. *Assume $|V| \leq M$ almost surely. Let $(\lambda_n)_{n\geq 1}$ be any deterministic sequence with*

$$\lambda_n \to \infty \qquad \textit{and} \qquad \sqrt{\lambda_n} = o(\log n), \tag{4}$$

e.g. $\lambda_n = (\log n)^{3/2}$. Define $\widehat{b}_n := -\frac{\log \widehat{L}_n(\lambda_n)}{\sqrt{2\lambda_n}}$. Then, $\widehat{b}_n \xrightarrow{P} b$.

Proof sketch: For a fixed value of λ, $\widehat{L}_n(\lambda)$ converges almost surely to $L_b(\lambda)$. The difficulty here is that λ_n increases with n. To deal with this problem, we show that, under condition (4), the Lindeberg–Feller Central Limit Theorem implies that $\sqrt{n}\cdot\frac{\widehat{L}_n(\lambda_n) - L_b(\lambda_n)}{\sqrt{\mathrm{Var}(\exp(-\lambda_n T))}}$ converges in distribution to $\mathcal{N}(0,1)$. From here, we establish that $\widehat{L}_n(\lambda_n)/L_b(\lambda_n) \xrightarrow{P} 1$. The full proof is provided in Appendix A.2. □

Richardson-extrapolated estimator. In our analysis of the estimator $\hat{b}_n$, the argument based on the central limit theorem required that $\sqrt{\lambda_n} = o(\log n)$. However, in our numerical experiments, this requirement leads to slow convergence rates. In our next result, a simple two-scale correction improves the *population bias*.

Theorem 3. *Assume $|V| \le M$ almost surely and that $\{\lambda_n\}_n$ satisfies* (4)*. Define*

$$\widetilde{B}_n \;:=\; \frac{\log \widehat{L}_n(\lambda_n) - \log \widehat{L}_n(4\lambda_n)}{\sqrt{2\lambda_n}}. \tag{5}$$

Then, $\widetilde{B}_n \xrightarrow{P} b$. Moreover, the population bias of $\widetilde{B}_n$ improves upon that of $\widehat{b}_n$ in the following sense:

$$\left| \frac{\log L_b(\lambda) - \log L_b(4\lambda)}{\sqrt{2\lambda}} - b \right| = O\!\left(\frac{1}{\lambda}\right), \qquad \left| -\frac{\log L_b(\lambda)}{\sqrt{2\lambda}} - b \right| = \Omega\!\left(\frac{1}{\sqrt{\lambda}}\right). \tag{6}$$

The full proof appears in Appendix A.3. The idea is based on the following simple observation. As shown in the proof, expanding $\log L_b(\lambda)$ as a function of λ and with $\lambda \to \infty$, yields

$$\log L_b(\lambda) \;=\; -b\sqrt{2\lambda} + \log(2C_0) - \frac{bC_2}{2C_0\sqrt{2\lambda}} + O(\lambda^{-1}), \tag{7}$$

where $C_0 := \mathbb{E}[\cosh(bV)]$ and $C_2 := \mathbb{E}[V^2 \cosh(bV)]$. Consequently, the one-scale estimator $\widehat{b}_n$ has an $O(\lambda_n^{-1/2})$ population bias, caused by the subleading term $\log(2C_0)$ in $\log L_b(\lambda)$. $\widetilde{B}_n$ then eliminates the dominating $O(\lambda_n^{-1/2})$ bias via a one-step Richardson extrapolation [59, 9].

3.3 Consistent estimation of the average drift without knowing the boundary

Returning to our main goal of estimating the utilitarian objective $\mu = \mathbb{E}_{V\sim F^*}[V]$, the natural next step is to plug in the estimator $\widetilde{B}_n$ into the estimator developed in Theorem 1, which assumed knowledge of the boundary b. Our next result shows that the resulting estimator is consistent. The proof is provided in Appendix A.4.

Theorem 4. *Assume that $|V| \le M$ almost surely and that $\{\lambda_n\}_n$ satisfies* (4)*. Let $\widehat{\mu}_n^{\mathrm{PI}} := \frac{1}{n}\sum_{i=1}^n z_i\, w_{\widetilde{B}_n}(t_i)$ where $\widetilde{B}_n$ is the estimator defined in Eq.* (5)*. Then, we have $\widehat{\mu}_n^{\mathrm{PI}} \xrightarrow{P} \mu$.*

Proof sketch: Decompose $\widehat{\mu}_n^{\mathrm{PI}} - \mu$ as $A_n + R_n$, with $A_n := \frac{1}{n}\sum_i z_i w_b(t_i) - \mu$ and $R_n := \frac{1}{n}\sum_i z_i\big(w_{\widetilde{B}_n}(t_i) - w_b(t_i)\big)$. We first note that $A_n \xrightarrow{P} 0$ by Theorem 1 and the law of large numbers. To show R_n also converges to zero in probability, we establish a envelope lemma (Lemma 2 in the appendix) that controls $|w_\beta(t)|$ by *gluing together three functions*: (i) $C_K(1 + t^{-1})$ as t is close to zero, (ii) a constant for t in a compact middle range, and (iii) an exponential bound $C_K e^{-c_K t}$ as t is large, all *uniformly* in K. Integrability of this envelope against the marginal law of t_i then yields $R_n \xrightarrow{P} 0$ by using the dominated convergence theorem together with the law of large numbers. □

Remark 2 (Sampling error and boundary estimation)**.** *Suppose $|V| \le M$ almost surely. The small-time decay of the first-passage density and the bounds in Lemma 2 imply $\sigma_{b,M}^2 := \sup_{|v|\le M} \mathbb{E}[w_b(T)^2 \mid V = v] < \infty$. Consequently, when b is known, unbiasedness and independence give $\mathbb{E}[(\widehat{\mu}_n - \mu)^2] \le \sigma_{b,M}^2/n$.*

To quantify the effect of estimating b, let $\widehat{b}$ be computed from an independent sample and let $K \subset (0,\infty)$ be a compact interval containing b in its interior. Differentiating the series for w_β and applying the small- and large-time estimates used in Lemma 2 gives $D_K^2 := \sup_{|v|\le M} \mathbb{E}\big[\sup_{\beta\in K} |\partial_\beta w_\beta(T)|^2 \,\big|\, V = v\big] < \infty$, where the conditional response-time law has boundary b. In particular, the derivative is $O(1 + t^{-1})$ near zero and is bounded by $C_K(1+t)e^{-c_K t}$ for large t. For n fresh observations, the mean-value theorem and the triangle inequality in L^2 therefore yield, on $\{\widehat{b} \in K\}$,

$$\left(\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n Z_i w_{\widehat{b}}(T_i) - \mu\right)^2 \middle| \widehat{b}\right]\right)^{1/2} \le \frac{\sigma_{b,M}}{\sqrt{n}} + D_K|\widehat{b} - b|. \tag{8}$$

For a consistent boundary estimator, this event has probability tending to one. The first term is the sampling error with a known boundary; the second propagates the boundary-estimation error. For the Richardson estimator, the latter includes the $O(\lambda_n^{-1})$ population bias in Theorem 3.

4 A Consistent Estimator for Heterogeneous Linear Preferences

We next move to the more general setting described in Section 2. We assume that, for each $x \in \mathcal{X}$, user i's utility is given by $u_i(x) = \phi(x)^\top \theta_i$ for some underlying latent preference vector $\theta_i \sim G^* \in \Delta(\mathbb{R}^d)$ and *shared* feature map $\phi : \mathcal{X} \to \mathbb{R}^d$. In this setting, given two alternatives x_i and y_i, the drift has the linear form $v_i = \psi_i^\top \theta_i$, where the parameter $\psi_i := \phi(x_i) - \phi(y_i) \in \mathbb{R}^d$ is known to the algorithm. Throughout this section, we assume that (ψ_i, θ_i) are exogenous and i.i.d. across i, and the DDM paths are independent. Our goal is to estimate $\theta^\star := \mathbb{E}_{\theta \sim G^*}[\theta]$.

We next show how the results derived in the previous section can be used to develop a consistent estimator of the mean preference vector $\theta^\star$. Let $\widehat{v}_i := z_i \, w_{\widetilde{B}_n}(t_i)$. Now consider the linear regression problem in which the i-th observation consists of the covariate vector ψ_i and the *pseudo-outcome* $\widehat{v}_i$. Define the sample covariance matrix $\widehat{Q}_n := \frac{1}{n}\sum_{i=1}^n \psi_i \psi_i^\top$ and empirical cross-moment $\widehat{m}_n := \frac{1}{n}\sum_{i=1}^n \psi_i \, \widehat{v}_i$. Then, the *plug-in regression estimator* is

$$\widehat{\theta}_n := \widehat{Q}_n^{-1} \widehat{m}_n. \tag{9}$$

We further assume that the covariance matrix $Q := \mathbb{E}[\psi_i \psi_i^\top]$ is positive definite and that the sample covariance matrix $\widehat{Q}_n$ is almost surely invertible, which holds when $n \geq d$ and the distribution of ψ_i's is non-degenerate [44].

Theorem 5. *Assume that there exist finite constants L and R such that $\|\psi_i\| \leq L$ and $\|\theta_i\| \leq R$ almost surely*[3]*, and that the sequence $\{\lambda_n\}_n$ satisfies* (4)*. Then, we have $\widetilde{B}_n \xrightarrow{P} b$ and $\widehat{\theta}_n \xrightarrow{P} \theta^\star$. Consequently, for every fixed $x \in \mathcal{X}$, $\phi(x)^\top \widehat{\theta}_n \xrightarrow{P} \phi(x)^\top \theta^\star$.*

Proof sketch: By the weak law of large numbers applied to $\widehat{Q}_n$, we have $\widehat{Q}_n \xrightarrow{P} Q$. Next, we establish that $\widehat{m}_n \xrightarrow{P} Q\theta^\star$. To do so, decompose $\widehat{m}_n = S_n + R_n$, where $S_n := \frac{1}{n}\sum_i \psi_i z_i w_b(t_i)$ and $R_n := \frac{1}{n}\sum_i \psi_i z_i (w_{\widetilde{B}_n}(t_i) - w_b(t_i))$. The term S_n has expectation $\mathbb{E}[\psi_i \, \mathbb{E}[z_i w_b(t_i) \mid \psi_i, \theta_i]] = \mathbb{E}[\psi_i \psi_i^\top \theta_i] = Q\theta^\star$ where the first equality follows from Theorem 1. Moreover, the residual term $R_n \xrightarrow{P} 0$ by the same argument of Theorem 4. Therefore, the continuous mapping theorem gives $\widehat{\theta}_n = \widehat{Q}_n^{-1}\widehat{m}_n \xrightarrow{P} Q^{-1}Q\theta^\star = \theta^\star$. The full proof is provided in Appendix A.5. □

Remark 3 (Identifiability with unknown boundary)**.** *Because the single-reward Bradley–Terry distribution depends on b and V only through their product bV, the mean preference vector $\theta^\star$ is* not *identifiable from choices alone when b is unknown—even asymptotically. However, as Theorem 5 shows, response times restore identifiability: the boundary and the utility scale are both pinned down by the introduced estimators.*

Remark 4 (Nonlinear population-average utilities)**.** *The response-time transformation applies to a general family of utilities $u(\cdot;\theta)$ with $\theta \sim G^*$. Let $\bar{u}(x) := \mathbb{E}_{\theta \sim G^*}[u(x;\theta)]$, and suppose that the labeler parameter is independent of the assigned comparison pair. For a known boundary, define $V_i^{\mathrm{RT}} := Z_i w_b(T_i)$. By Theorem 1, $\mathbb{E}[V_i^{\mathrm{RT}} \mid x_i, y_i, \theta_i] = u(x_i;\theta_i) - u(y_i;\theta_i)$, $\quad \mathbb{E}[V_i^{\mathrm{RT}} \mid x_i, y_i] = \bar{u}(x_i) - \bar{u}(y_i)$. Suppose $\bar{u} = h_{\eta^\star}$ for a function class $\{h_\eta : \eta \in \mathcal{H}\}$, which may be a nonlinear parametric class or a neural-network class. Writing $\Delta_\eta(x,y) := h_\eta(x) - h_\eta(y)$ gives the empirical objective*

$$\widehat{\eta}_n \in \arg\min_{\eta \in \mathcal{H}} \frac{1}{n}\sum_{i=1}^n \big(\Delta_\eta(x_i, y_i) - Z_i w_b(T_i)\big)^2. \tag{10}$$

If $\mathbb{E}[(V^{\mathrm{RT}})^2] < \infty$, its population risk satisfies

$$\mathcal{R}(\eta) := \mathbb{E}[(\Delta_\eta(X,Y) - V^{\mathrm{RT}})^2] = \mathbb{E}[(\Delta_\eta(X,Y) - \Delta_{\eta^\star}(X,Y))^2] + \mathbb{E}[\mathrm{Var}(V^{\mathrm{RT}} \mid X, Y)].$$

Hence $\mathcal{R}(\eta) - \mathcal{R}(\eta^\star) = \mathbb{E}[(\Delta_\eta - \Delta_{\eta^\star})^2] \geq 0$: the squared loss is Fisher consistent for the population-average utility differences. Therefore, $\eta^\star$ is the global optimizer of the population risk and consistency is given by uniform law of large numbers for the loss function class.

[3] We relax these assumptions in Appendix B.3, thus accommodating bounded ψ_i and Gaussian θ_i.

4.1 Recovering the utilitarian aggregator

Consider a utilitarian planner who wants to choose the alternative (identified by its feature vector in $\mathbb{R}^d$) that maximizes society's average utility, given by[4]

$$\mathrm{AvgUtil}(x) = \mathbb{E}_{\theta\sim G^*}[x^\top\theta] = x^\top\theta^\star. \tag{11}$$

Assume $\sup_{x\in\mathcal{X}} \mathrm{AvgUtil}(x) > 0$. The planner has access to a dataset $\mathcal{D}_n$ consisting of n choice observations, possibly augmented with response-time observations, from n users. The planner then applies a possibly randomized algorithm $\mathcal{A}$ to $\mathcal{D}_n$. The expected average utility of $\mathcal{A}$, denoted by $\mathrm{AvgUtil}_n(\mathcal{A})$, is defined as $\mathrm{AvgUtil}_n(\mathcal{A}) := \mathbb{E}\big[\mathrm{AvgUtil}\big(\mathcal{A}(\mathcal{D}_n)\big)\big]$, where the expectation is taken over the randomness in the dataset $\mathcal{D}_n$ and any additional randomness in the algorithm.

The distortion of $\mathcal{A}$ is the worst-case asymptotic competitive ratio between the optimal average utility and the average utility achieved by the algorithm. This is formally defined as:

$$\mathrm{dist}(\mathcal{A}) := \sup_G \limsup_{n\to\infty} \frac{\sup_{x\in\mathcal{X}} \mathrm{AvgUtil}(x)}{\mathrm{AvgUtil}_n(\mathcal{A})}. \tag{12}$$

In other words, the distortion quantifies whether there is an inherent limit on how well a utilitarian decision maker can recover the optimal decision from preference data, even as the dataset size $\to \infty$.

Let us first consider the setting without response-time data. Gölz et al. [28, Theorem 3] provide a lower bound of $\frac{\beta}{2}\cdot\frac{1+\exp(-\beta)}{1-\exp(-\beta)}$ on the distortion rate, where β is the temperature parameter, corresponding to the boundary b in our model up to a constant factor of 2. This lower bound can also be realized by a linear-utility model in higher dimensions. Thus, even with infinitely many choice observations, a utilitarian planner may fail to recover an alternative with optimal average utility.

Appendix B.4 gives a formal statement and proof of this finite-dimensional construction.

On the other hand, suppose that response-time data are also included in the dataset $\mathcal{D}$. A natural algorithm $\mathcal{A}$ in this setting is to select $\arg\max_{x\in\mathcal{X}} x^\top\widehat{\theta}_n$. Assume that the alternatives are uniformly bounded, and that the above maximizer exists. Then Theorem 5 implies that, for every $\varepsilon > 0$, the probability that the gap between the maximum average utility and the average utility of the algorithm's output exceeds ε converges to zero. Consequently, the distortion rate in this setting is equal to one. In this sense, response-time data allow the utilitarian planner to bypass the choice-only distortion barrier and asymptotically recover an optimal decision.

5 Experiments

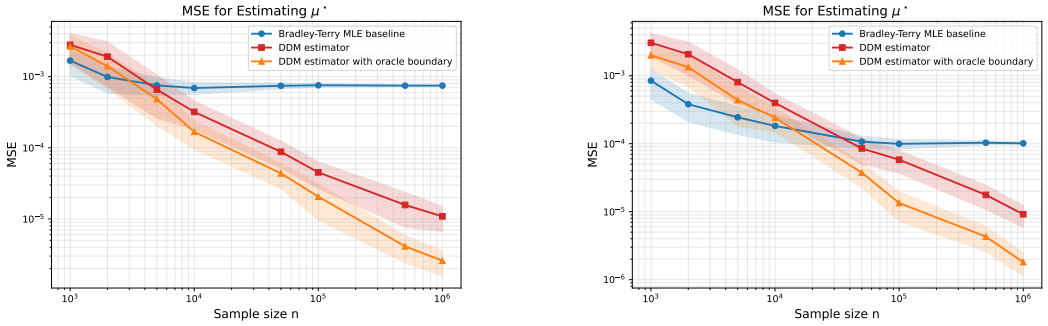
We validate our estimators from Sections 3 and 4 on synthetic DDM data and on an intertemporal-choice dataset, collected by Amasino et al. [3]. In every experiment the boundary is treated as *unknown* and estimated from response times alone via the bias-corrected estimator $\widetilde{B}_n$ of Eq. (5) with $\lambda_n = (\log n)^{3/2}$; the weight w_β is truncated at $K = 100$ terms; and each reported value averages over $R = 50$ independent Monte-Carlo replications. Full experimental details appear in Appendix C.

5.1 Synthetic experiments

Data generating process. We study two regimes mirroring Sections 3 and 4. In the *tabular* setting, we fix two alternatives and sample drifts $v_1, \ldots, v_n$ i.i.d. from F^* with mean $\mu^\star = 0.25$ and boundary $b^\star = 1.25$; we consider both Uniform and Beta priors on the drift to reflect heterogeneous preference. In the *linear* setting, we consider the linear drift model $v_i = \psi_i^\top\theta_i$ with unknown boundary $b = 1.25$. Each alternative vector $\psi_i \sim \mathrm{Uniform}[-1,1]^4$, and the mean preference feature is $\theta^\star = (0.25, -0.15, 0.10, -0.30)$, where each coordinate of θ_i follows one of four priors with variance $\sigma_\theta^2 = 0.25$: Gaussian, Uniform, Beta, or Laplace (truncated to $\pm 6\sigma_\theta$).

Methods. In both regimes we compare the following three estimators.

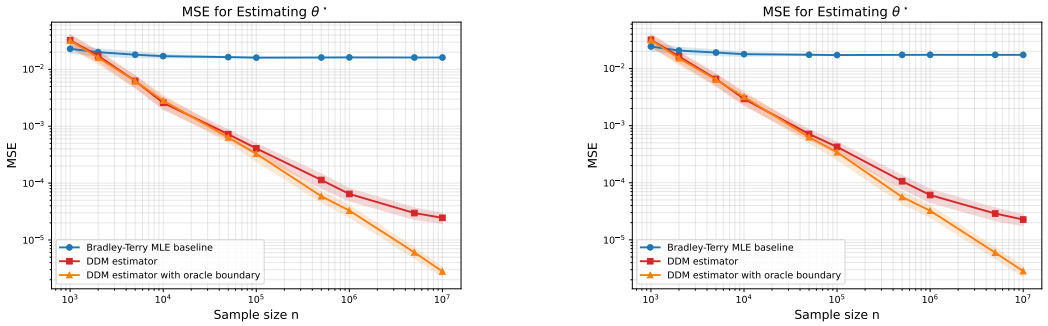
[4]For any distribution $\kappa \in \Delta(\mathcal{X})$, we similarly define $\mathrm{AvgUtil}(\kappa) := \mathbb{E}_{x\sim\kappa}[\mathrm{AvgUtil}(x)]$.



(a) Uniform prior on V.

(b) Beta prior on V.

Figure 1: **Tabular setting.** Monte-Carlo MSE of $\widehat{\mu}_n$ versus sample size n for two latent-drift priors. The DDM estimator dominates the baseline, which plateaus at a heterogeneity-induced bias floor. Shaded bands denote pointwise 95% confidence intervals over $R = 50$ replications.



(a) Gaussian.

(b) Uniform.

Figure 2: **Linear contextual setting.** Monte-Carlo MSE of $\widehat{\theta}_n$ versus sample size n for two priors on $\theta_i \in \mathbb{R}^4$. The DDM estimator again tracks the oracle and uniformly dominates the baseline. Shaded bands denote pointwise 95% confidence intervals over $R = 50$ replications.

- *Bradley–Terry MLE baseline*: the MLE minimizer of Bradley–Terry (temperature calibrated by the oracle boundary $b^\star$) over the choice-only data. In the *tabular* setting, $\widehat{\mu}_n^{\mathrm{BT}} = \operatorname{arctanh}\big(\frac{1}{n}\sum_i Z_i\big)/b^\star$; in the *linear* setting, the pooled Bradley–Terry logistic MLE $\widehat{\theta}_n^{\mathrm{BT}} = \arg\min_\theta \frac{1}{n}\sum_i \log(1 + e^{-2b^\star Z_i \psi_i^\top \theta})$.
- *Drift Diffusion Model (DDM)*: the consistent estimator developed by Theorem 4 and Theorem 5. In the *tabular* setting, $\hat{\mu}_n^{\mathrm{DDM}} = \frac{1}{n}\sum_i Z_i w_{\widetilde{B}_n}(\tau_i)$; in the *linear* setting, $\hat{\theta}_n^{\mathrm{DDM}}$ is the OLS estimator in Eq. (9). Both use the bias-corrected boundary estimate $\widetilde{B}_n$.
- *DDM with Oracle Boundary:* identical to the previous Drift Diffusion Model estimator except replaces $\widetilde{B}_n$ by the oracle boundary $b^\star$.

Results. We measure the Monte-Carlo MSE to the ground-truth population target: $\mathbb{E}[(\widehat{\mu}_n - \mu^\star)^2]$ in the tabular case and $\mathbb{E}[\|\widehat{\theta}_n - \theta^\star\|_2^2]$ in the linear case. Figures 1 and 2 show two robust patterns. First, the Bradley–Terry baseline plateaus at a nonzero bias floor. In comparison, the error of our DDM estimator decays as sample size grows, and achieves substantially smaller errors. Second, using the estimated boundary $\widetilde{B}_n$ incurs only moderate degradation relative to the oracle-boundary estimator, supporting the practical effectiveness of the Richardson correction.

5.2 Real-data experiment

Setup. We use the publicly available intertemporal-choice dataset of Amasino et al. [3]. Each trial asks a participant to choose between a smaller-sooner reward $s_r \in \{0.5, \ldots, 9.5\}$ paid today

and a larger-later reward $\ell_r = 10$ paid after delay $\ell_d \in \{1, 7, 15, 30, 90, 180, 365\}$ days, with response time τ_i recorded. Following Echenique et al. [19], we encode the trial by $\psi_i = \big((\ell_r - s_r)/9.5,\ -\ell_d/365\big) \in \mathbb{R}^2$. After dropping missing trials and two participants with only one choice class, the pooled sample contains $n_{\max} = 13{,}793$ observations from $S = 98$ participants. The target is the subject-level average preference feature $\theta^\star = \frac{1}{S}\sum_{s=1}^{S} \widehat{\theta}^{(s)} \in \mathbb{R}^2$, computed by averaging per-participant Bradley–Terry MLE $\widehat{\theta}^{(s)}$. Then we subsample n anonymous data points used to estimate $\theta^\star$.

Methods. We compare the following two estimators.

- *Bradley–Terry MLE baseline*: the Bradley–Terry logistic MLE over choice-only data.
- *Drift Diffusion Model (DDM)*: the OLS estimator in Eq. (9) using the bias-corrected boundary estimate $\widetilde{B}_n$.

Results. Because the common boundary is unknown, only the *direction* of $\theta^\star$ is comparable across estimators. We therefore report $\mathrm{CosSim}(\widehat{\theta}, \theta^\star)$, averaged over $R = 50$ subsamples of size n. Figure 3 shows that the DDM estimator is closer to the target than the Bradley–Terry MLE baseline at nearly every subsample sizes. As n grows, the baseline plateaus around $\mathrm{CosSim} \approx 0.993 - 0.994$, whereas our method reaches about 0.998 at the largest evaluated subsample size $n = 5000$. This is consistent with the theory: response times retain preference information that choices alone discard, even when the DDM is only an approximation.

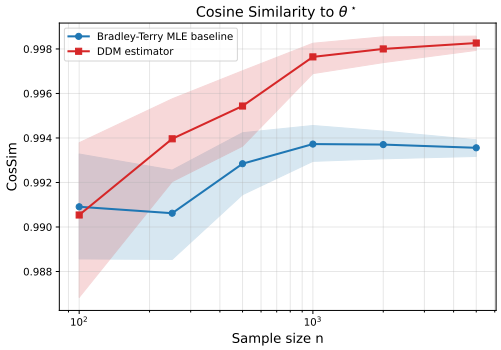


Figure 3: **Intertemporal choice data.** Cosine similarity between each estimator and the subject-level target versus subsample size n on Amasino et al. [3] dataset. Shaded regions are 95% bootstrap bands over $R = 50$ replications. Our DDM estimator achieves higher accuracy than the baseline.

6 Conclusion

Under a drift-diffusion model with a common but unknown decision boundary, binary choices combined with response times are sufficient to construct a consistent estimator of the population-average preference in a heterogeneous anonymous labeler pool, a target that is not identifiable from choices alone.

Broader impact. Response times are currently not part of the standard preference-learning datasets for LLM alignment, even though response time may be collected essentially at zero cost. Our paper shows that logging these timestamps in future datasets can mitigate the Bradley–Terry distortion that arises when aligning LLMs to heterogeneous feedback, without the need to collect extra labels or stable annotator identifiers. At the same time, timing metadata can itself be sensitive or correlated with protected attributes, so any deployment should treat response times as potentially privacy-relevant signals rather than frictionless exhaust.

Limitations and future work. The theory assumes that choices are well-described by a DDM with a common boundary; in practice, caution may vary across annotators or over time, and the model can be misspecified in other ways as well. The relaxation of these assumptions is discussed in the appendix. Moreover, the boundary estimator relies on a large-λ Laplace tail and converges more slowly in finite samples than the drift estimator. Extending the analysis to heterogeneous boundaries and integrating the response-time pseudo-outcomes into end-to-end alignment objectives are natural next steps.

Acknowledgment

Part of this work was supported by the Simons Institute for the Theory of Computing, and conducted when Alireza Fallah visited the Institute. We also wish to acknowledge funding by the European Union (ERC-2022-SYG-OCEAN-101071601). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] H. Alipour and M. Goyal. Utilitarian distortion under probabilistic voting. *arXiv preprint arXiv:2602.11152*, 2026. URL `https://arxiv.org/abs/2602.11152`.

[2] C. Alós-Ferrer, E. Fehr, and N. Netzer. Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy*, 129(6):1828–1877, 2021.

[3] D. R. Amasino, N. J. Sullivan, R. E. Kranton, and S. A. Huettel. Amount and time exert independent influences on intertemporal choice. *Nature Human Behaviour*, 3(4):383–392, 2019.

[4] Anthropic. The Claude 3 model family: Opus, Sonnet, Haiku. Model card, Anthropic, 2024.

[5] M. G. Azar, Z. D. Guo, B. Piot, R. Munos, M. Rowland, M. Valko, and D. Calandriello. A general theoretical paradigm to understand learning from human preferences. In *International Conference on Artificial Intelligence and Statistics*, pages 4447–4455. PMLR, 2024.

[6] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

[7] P. Billingsley. *Probability and Measure*. John Wiley & Sons, 2017.

[8] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[9] C. Brezinski and M. Redivo-Zaglia. *Extrapolation Methods: Theory and Practice*. North-Holland, 1991.

[10] C. Calluso, A. Tosoni, L. Cannito, and G. Committeri. Concreteness and emotional valence of episodic future thinking (EFT) independently affect the dynamics of intertemporal decisions. *PLOS ONE*, 14(5):e0217224, 2019. doi: 10.1371/journal.pone.0217224.

[11] J. D. Chang, W. Zhan, O. Oertell, K. Brantley, D. Misra, J. D. Lee, and W. Sun. Dataset reset policy optimization for RLHF. *arXiv preprint arXiv:2404.08495*, 2024.

[12] K. Chidambaram, K. V. Seetharaman, and V. Syrgkanis. Direct preference optimization with unobserved preference heterogeneity. *arXiv preprint arXiv:2405.15065*, 2024. URL `https://arxiv.org/abs/2405.15065`.

[13] K. X. Chiong, M. Shum, R. Webb, and R. Chen. Combining choice and response time data: A drift-diffusion model of mobile advertisements. *Management Science*, 70(2):1238–1257, 2024.

[14] S. R. Chowdhury, A. Kini, and N. Natarajan. Provably robust DPO: Aligning language models with noisy feedback. *arXiv preprint arXiv:2403.00409*, 2024.

[15] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei. Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems*, 30, 2017.

[16] J. A. Clithero. Improving out-of-sample predictions using response times and a model of the decision process. *Journal of Economic Behavior & Organization*, 148:344–375, 2018.

[17] A. Copeland, T. Stafford, and M. Field. Methodological issues with value-based decision-making (VBDM) tasks: The effect of trial wording on evidence accumulation outputs from the EZ drift-diffusion model. *Cogent Psychology*, 9(1):2079801, 2022. doi: 10.1080/23311908.2022.2079801.

[18] C. d'Aspremont and L. Gevers. Equity and the informational basis of collective choice. *The Review of Economic Studies*, 44(2):199–209, 1977.

[19] F. Echenique, A. Fallah, and M. I. Jordan. A general framework for estimating preferences using response time data, 2025. arXiv preprint arXiv:2507.20403.

[20] J. Eisenstein, C. Nagpal, A. Agarwal, A. Beirami, A. D'Amour, D. Dvijotham, A. Fisch, K. Heller, S. Pfohl, D. Ramachandran, et al. Helping or herding? reward model ensembles mitigate but do not eliminate reward hacking. *arXiv preprint arXiv:2312.09244*, 2023.

[21] K. Ethayarajh, W. Xu, N. Muennighoff, D. Jurafsky, and D. Kiela. KTO: Model alignment as prospect theoretic optimization. *arXiv preprint arXiv:2402.01306*, 2024.

[22] E. Fehr and A. Rangel. Neuroeconomic foundations of economic choice—recent advances. *Journal of Economic Perspectives*, 25(4):3–30, 2011.

[23] D. Fudenberg, P. Strack, and T. Strzalecki. Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–3684, 2018.

[24] D. Fudenberg, W. K. Newey, P. Strack, and T. Strzalecki. Testing the drift-diffusion model. *Proceedings of the National Academy of Sciences*, 117(52):33141–33148, 2020.

[25] L. Gao, J. Schulman, and J. Hilton. Scaling laws for reward model overoptimization. In *International Conference on Machine Learning*, pages 10835–10866. PMLR, 2023.

[26] Gemini Team, Google. Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023. URL `https://arxiv.org/abs/2312.11805`.

[27] J. I. Gold and M. N. Shadlen. The neural basis of decision making. *Annual Review of Neuroscience*, 30:535–574, 2007.

[28] P. Gölz, N. Haghtalab, and K. Yang. Distortion of AI alignment: Does preference optimization optimize for preferences? In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, pages 25969–26007. Curran Associates, Inc., 2025.

[29] M. Goyal and S. Sarmasarkar. Metric distortion under probabilistic voting. In *Proceedings of the 26th ACM Conference on Economics and Computation*, 2025. URL `https://arxiv.org/abs/2405.14223`.

[30] J. A. Grange. Computational modelling of the speed–accuracy tradeoff: No evidence for an association with depression symptomatology. *Journal of Psychiatric Research*, 147:111–125, 2022. doi: 10.1016/j.jpsychires.2021.12.057.

[31] J. Hong, N. Lee, and J. Thorne. ORPO: Monolithic preference optimization without reference model. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 11170–11189, 2024. URL `https://arxiv.org/abs/2403.07691`.

[32] S. A. Jones, J. S. Steele, and B. J. Nagel. Binge drinking and family history of alcoholism are associated with an altered developmental trajectory of impulsive choice across adolescence. *Addiction*, 112(7):1184–1192, 2017. doi: 10.1111/add.13823.

[33] J. W. Kable, M. K. Caulfield, M. Falcone, M. McConnell, L. Bernardo, T. Parthasarathi, N. Cooper, R. Ashare, J. Audrain-McGovern, R. Hornik, P. Diefenbach, F. J. Lee, and C. Lerman. No effect of commercial cognitive training on brain activity, choice behavior, or cognitive performance. *Journal of Neuroscience*, 37(31):7390–7402, 2017. doi: 10.1523/JNEUROSCI.2832-16.2017.

[34] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer Science & Business Media, 1991.

[35] I. Krajbich, C. Armel, and A. Rangel. Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13(10):1292–1298, 2010.

[36] K. Kuroda, A. N. Tump, and R. H. J. M. Kurvers. Individual differences in speed–accuracy trade-off influence social decision-making in dyads. *Proceedings of the Royal Society B: Biological Sciences*, 292(2051):20251077, 2025. doi: 10.1098/rspb.2025.1077.

[37] F. Lange and F. Eggert. Sweet delusion. glucose drinks fail to counteract ego depletion. *Appetite*, 75:54–63, 2014. doi: 10.1016/j.appet.2013.12.020.

[38] S. Li, Y. Zhang, Z. Ren, C. Liang, N. Li, and J. A. Shah. Enhancing preference-based linear bandits via human response time. *Advances in Neural Information Processing Systems*, 37: 16852–16893, 2024.

[39] Y. Li and J. L. McClelland. A weighted constraint satisfaction approach to human goal-directed decision making. *PLOS Computational Biology*, 18(6):e1009553, 2022. doi: 10.1371/journal.pcbi.1009553.

[40] K. Liu, Q. Long, Z. Shi, W. J. Su, and J. Xiao. Statistical impossibility and possibility of aligning LLMs with human preferences: From Condorcet paradox to Nash equilibrium. *arXiv preprint arXiv:2503.10990*, 2025. URL `https://arxiv.org/abs/2503.10990`.

[41] Z. Liu, M. Lu, S. Zhang, B. Liu, H. Guo, Y. Yang, J. Blanchet, and Z. Wang. Provably mitigating overoptimization in RLHF: Your SFT loss is implicitly an adversarial regularizer. *Advances in Neural Information Processing Systems*, 37:138663–138697, 2024.

[42] Llama Team, AI @ Meta. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024. URL `https://arxiv.org/abs/2407.21783`.

[43] Y. Meng, M. Xia, and D. Chen. SimPO: Simple preference optimization with a reference-free reward. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. URL `https://arxiv.org/abs/2405.14734`.

[44] J. Mourtada. Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices. *The Annals of Statistics*, 50(4):2157–2178, 2022.

[45] R. Munos, M. Valko, D. Calandriello, M. G. Azar, M. Rowland, Z. D. Guo, Y. Tang, M. Geist, T. Mesnard, A. Michi, et al. Nash learning from human feedback. *arXiv preprint arXiv:2312.00886*, 2023. URL `https://arxiv.org/abs/2312.00886`.

[46] OpenAI. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023. URL `https://arxiv.org/abs/2303.08774`.

[47] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

[48] A. Ozkan and J. Zhang. Information sources and congruency modulate preference-based decision-making processes. *Journal of Cognitive Psychology*, 36(6):775–792, 2024. doi: 10.1080/20445911.2024.2384666.

[49] C. Park, M. Liu, D. Kong, K. Zhang, and A. Ozdaglar. RLHF from Heterogeneous Feedback via Personalization and Preference Aggregation. *arXiv preprint arXiv:2405.00254*, 2024. URL `https://arxiv.org/abs/2405.00254`.

[50] H. Pongratz and M. Schoemann. A large-scale dataset of choice and response-time data in intertemporal choice. *Scientific Data*, 13:323, 2026. doi: 10.1038/s41597-026-06947-4.

[51] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn. Direct preference optimization: Your language model is secretly a reward model. *arXiv preprint arXiv:2305.18290*, 2023. URL `https://arxiv.org/abs/2305.18290`.

[52] R. Ratcliff. A theory of memory retrieval. *Psychological Review*, 85(2):59–108, 1978.

[53] R. Ratcliff. Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3): 510–532, 1993. doi: 10.1037/0033-2909.114.3.510.

[54] R. Ratcliff and G. McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922, 2008.

[55] R. Ratcliff and J. N. Rouder. Modeling response times for two-choice decisions. *Psychological Science*, 9(5):347–356, 1998.

[56] R. Ratcliff and F. Tuerlinckx. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, 9(3):438–481, 2002.

[57] R. Ratcliff, P. L. Smith, S. D. Brown, and G. McKoon. Diffusion decision model: current issues and history. *Trends in Cognitive Sciences*, 20(4):260–281, 2016.

[58] L. M. Reich, L. A. M. Mahr, M. Vacondio, and A. S. Khalid. The effect of ingroup vs. outgroup members' behavior on charity preference: A drift-diffusion model approach. *Frontiers in Psychology*, 13:854747, 2022. doi: 10.3389/fpsyg.2022.854747.

[59] L. F. Richardson. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society of London. Series A*, 210:307–357, 1911.

[60] A. Sawarni, S. Sarmasarkar, and V. Syrgkanis. Preference learning with response time: Robust losses and guarantees. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, pages 104110–104156. Curran Associates, Inc., 2025.

[61] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[62] D. Schwenke, P. Wehner, and S. Scherbaum. Effects of individual and dyadic decision-making and normative reference on delay discounting decisions. *Cognitive Research: Principles and Implications*, 7(1):71, 2022. doi: 10.1186/s41235-022-00422-5.

[63] Z. Shi, K. Liu, Q. Long, W. J. Su, and J. Xiao. Fundamental limits of game-theoretic LLM alignment: Smith consistency and preference matching. *arXiv preprint arXiv:2505.20627*, 2025. URL `https://arxiv.org/abs/2505.20627`.

[64] A. Shirali, A. Nasr-Esfahany, A. Alomar, P. Mirtaheri, R. Abebe, and A. Procaccia. Direct alignment with heterogeneous preferences. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, pages 69858–69898. Curran Associates, Inc., 2025.

[65] A. Siththaranjan, C. Laidlaw, and D. Hadfield-Menell. Distributional preference learning: Understanding and accounting for hidden context in RLHF. *International Conference on Learning Representations (ICLR)*, 2024. URL `https://arxiv.org/abs/2312.08358`.

[66] T. Sorensen, L. Jiang, J. Hwang, S. Levine, V. Pyatkin, P. West, N. Dziri, X. Lu, K. Rao, C. Bhagavatula, M. Sap, J. Tasioulas, and Y. Choi. Value Kaleidoscope: Engaging AI with Pluralistic Human Values, Rights, and Duties. *arXiv preprint arXiv:2309.00779*, 2023. URL `https://arxiv.org/abs/2309.00779`.

[67] N. Stiennon, L. Ouyang, J. Wu, D. Ziegler, R. Lowe, C. Voss, A. Radford, D. Amodei, and P. F. Christiano. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems*, 33:3008–3021, 2020.

[68] T. Strzalecki. Stochastic choice theory. *Cambridge University Press*, 2025.

[69] Y. Tang, D. Z. Guo, Z. Zheng, D. Calandriello, Y. Cao, E. Tarassov, R. Munos, B. Á. Pires, M. Valko, Y. Cheng, et al. Understanding the performance gap between online and offline alignment algorithms. *arXiv preprint arXiv:2405.08448*, 2024.

[70] Y. Tang, Z. D. Guo, Z. Zheng, D. Calandriello, R. Munos, M. Rowland, P. H. Richemond, M. Valko, B. Á. Pires, and B. Piot. Generalized preference optimization: A unified approach to offline alignment. *arXiv preprint arXiv:2402.05749*, 2024.

[71] A. Weigard, M. F. Molloy, C. Sripada, and A. Heathcote. The diffusion model's drift rate parameter primarily reflects efficiency, rather than speed, of evidence accumulation. *Psychonomic Bulletin and Review*, 33:100, 2026. doi: 10.3758/s13423-026-02861-3.

[72] J. Xiao, Z. Shi, K. Liu, Q. Long, and W. J. Su. Theoretical tensions in RLHF: Reconciling empirical success with inconsistencies in social choice theory. *arXiv preprint arXiv:2506.12350*, 2025. URL `https://arxiv.org/abs/2506.12350`.

[73] X.-L. Yang, S.-T. Chen, and H.-Z. Liu. The effect of incentives on intertemporal choice: Choice, confidence, and eye movements. *Frontiers in Psychology*, 13:989511, 2022. doi: 10.3389/fpsyg.2022.989511.

[74] Y. Zhao, R. Joshi, T. Liu, M. Khalman, M. Saleh, and P. J. Liu. SLiC-HF: Sequence likelihood calibration with human feedback. *arXiv preprint arXiv:2305.10425*, 2023.

[75] B. Zhu, J. Jiao, and M. I. Jordan. Principled reinforcement learning with human feedback from pairwise or k-wise comparisons. In *International Conference on Machine Learning*. PMLR, 2023.

[76] B. Zhu, H. Sharma, F. V. Frujeri, S. Dong, C. Zhu, M. I. Jordan, and J. Jiao. Fine-tuning language models with advantage-induced policy alignment. In *Proceedings of the 59th Annual Allerton Conference on Communication, Control, and Computing*, 2023.

[77] D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019.

A Omitted Proofs

This appendix collects the full proofs of the results stated in the main body.

A.1 Proof of Theorem 1

Fix $b > 0$. We first introduce some notation. Let $p_v(z,t)$ denote the joint density of (Z,T) under drift $V = v$. Let $\mathbb{P}_v$ denote the marginal distribution of Z. Finally, $\mathbb{E}_v$ denotes the expectation under a DDM with fixed drift $V = v$ and boundary b. The proof has three stages.

Stage 1: Joint distribution of Z and T. Applying Girsanov's theorem, we obtain the tilting identity (see also [19][Lemma 3])

$$p_v(z,t) \;=\; p_0(z,t)\,\exp\!\big(bvz - \frac{1}{2}v^2 t\big).$$

Under zero drift and symmetric boundaries, Z is symmetric and independent of T, so $p_0(z,t) = \frac{1}{2}f_0(t)$ for some density f_0. Hence

$$p_v(z,t) \;=\; \frac{1}{2}f_0(t)\,\exp\!\big(bvz - \frac{1}{2}v^2 t\big). \tag{13}$$

Summing (13) over $z \in \{\pm 1\}$ gives the marginal of T under drift v, and the conditional probability of Z given T is

$$\mathbb{P}_v(Z = +1 \mid T = t) \;=\; \frac{e^{bv}}{e^{bv} + e^{-bv}} \;=\; \frac{1}{1 + e^{-2bv}},$$

which does not depend on t. Therefore $Z \perp\!\!\!\perp T$ under drift v, with $\mathbb{E}_v[Z] = \tanh(bv)$.

Stage 2: Reduction to a Laplace identity. Because Z and T are independent under drift v,

$$\mathbb{E}_v[Z w_b(T)] \;=\; \tanh(bv)\,\mathbb{E}_v[w_b(T)]. \tag{14}$$

Therefore, to establish $\mathbb{E}_v[Z w_b(T)] = v$ for all v, it suffices to show

$$\mathbb{E}_v[w_b(T)] \;=\; \frac{v}{\tanh(bv)}. \tag{15}$$

Using (13),

$$\mathbb{E}_v[w_b(T)] = \int_0^\infty w_b(t) f_0(t)\,\frac{e^{bv} + e^{-bv}}{2}\,e^{-v^2 t/2}\,dt = \cosh(bv)\,G(v^2/2),$$

where $G(s) := \int_0^\infty w_b(t) f_0(t) e^{-st}\,dt$. Note that $G(s)$ is the Laplace transform of $w_b(t)f_0(t)$. Matching to (15) and canceling $\cosh(bv)$, we can see that we have to show

$$G(s) \;=\; \frac{\sqrt{2s}}{\sinh(b\sqrt{2s})}. \tag{16}$$

In other words, we have to show that $w_b(t)f_0(t)$ is the inverse Laplace transform of the right-hand side.

Stage 3: Series inversion. First note that the distribution of response time under zero drift, i.e., $f_0(t)$, is given by [34][Equation 8.24]

$$f_0(t) \;=\; \sqrt{\frac{2}{\pi}}\,t^{-3/2}\sum_{k=0}^{\infty}(-1)^k c_k(b)\exp\!\left(-\frac{c_k(b)^2}{2t}\right). \tag{17}$$

Next, note that, by expanding $1/\sinh(y) = 2e^{-y}/(1 - e^{-2y}) = 2\sum_{k\geq 0} e^{-(2k+1)y}$ and multiplying by $\sqrt{2s}$, we have

$$\frac{\sqrt{2s}}{\sinh(b\sqrt{2s})} \;=\; 2\sum_{k=0}^{\infty}\sqrt{2s}\,e^{-c_k(b)\sqrt{2s}}.$$

Now, to compute the inverse Laplace transform of the left hand side, first note that the inverse Laplace transform is a linear operator, and so, it suffices to find the inverse Laplace transform of $\mathcal{L}^{-1}[\sqrt{2s}\, e^{-c\sqrt{2s}}]$ for every c.

Notice that

$$\mathcal{L}^{-1}\left[e^{-c\sqrt{s}}\right](t) = \frac{c}{2\sqrt{\pi}\, t^{3/2}} \exp\left(-\frac{c^2}{4t}\right), \qquad c > 0,\ t > 0. \tag{18}$$

Next, differentiating $\mathcal{L}^{-1}[e^{-c\sqrt{2s}}](t)$ with respect to c gives $-\mathcal{L}^{-1}[\sqrt{2s}\, e^{-c\sqrt{2s}}](t) = \frac{1}{\sqrt{2\pi}} t^{-3/2} e^{-c^2/(2t)}(\frac{c^2}{t} - 1)$.

Therefore, by linearity of the inverse Laplace transform, we have

$$q(t) := \mathcal{L}^{-1}\left[\frac{\sqrt{2s}}{\sinh(b\sqrt{2s})}\right](t) \;=\; \sqrt{\frac{2}{\pi}}\, t^{-3/2} \sum_{k=0}^{\infty} \exp\left(-\frac{c_k(b)^2}{2t}\right)\left(\frac{c_k(b)^2}{t} - 1\right). \tag{19}$$

Recall that our goal was to show the $q(t)$ above is equal to $f_0(t)w_b(t)$. Dividing (19) by (17) and canceling the common prefactor $\sqrt{2/\pi}\, t^{-3/2}$ implies that we have to show $w_b(t)$ is equal to

$$\frac{q(t)}{f_0(t)} \;=\; \frac{\sum_{k\geq 0} e^{-c_k(b)^2/(2t)}(c_k(b)^2/t - 1)}{\sum_{k\geq 0}(-1)^k c_k(b) e^{-c_k(b)^2/(2t)}},$$

which is exactly (2). This gives us the desired result.

Stage 4: Unbiasedness and minimality. By (14) and (15), $\mathbb{E}_v[Zw_b(T)] = v$ for every v. Iterating expectation under the mixture $V \sim F^*$ gives $\mathbb{E}[Zw_b(T)] = \mathbb{E}[V] = \mu$, so $\mathbb{E}[\widehat{\mu}_n] = \mu$.

Finally, uniqueness of w_b follows from the injectivity of the Laplace transform: if another smooth $\tilde{w}$ satisfied (15), then $\tilde{w} \cdot f_0$ and $w_b \cdot f_0$ would share the transform in (16), and injectivity would force $\tilde{w} = w_b$ Lebesgue-almost everywhere, hence everywhere by continuity. □

A.2 Proof of Theorem 2

With a slight abuse of notation, let $L_b(\lambda) := \mathbb{E}[e^{-\lambda T}]$, where the expectation is taken over the randomness of both V and T.

Step 1: Population tail rate. Fix $b > 0$ and $\lambda \geq 0$. Under drift v, as stated in (3), $\mathbb{E}[e^{-\lambda T} \mid V = v] = \cosh(bv)/\cosh(b\sqrt{v^2 + 2\lambda})$; therefore, we have:

$$L_b(\lambda) \;=\; \mathbb{E}_{V\sim F^*}\left[\frac{\cosh(bV)}{\cosh(b\sqrt{V^2 + 2\lambda})}\right]. \tag{20}$$

For $v \in [-M, M]$, $\cosh(bv) \geq 1$ and $\cosh(x) \leq e^x$, so $\cosh(bv)/\cosh(b\sqrt{v^2 + 2\lambda}) \geq e^{-b\sqrt{M^2+2\lambda}}$, yielding

$$L_b(\lambda) \;\geq\; e^{-b\sqrt{M^2+2\lambda}}. \tag{21}$$

Conversely, $\cosh(bv) \leq e^{bM}$ and $\cosh(x) \geq e^x/2$, so $\cosh(bv)/\cosh(b\sqrt{v^2 + 2\lambda}) \leq 2e^{bM}e^{-b\sqrt{2\lambda}}$, giving

$$L_b(\lambda) \;\leq\; 2e^{bM}e^{-b\sqrt{2\lambda}}. \tag{22}$$

Combining (21) and (22) implies

$$-\frac{b\sqrt{M^2 + 2\lambda}}{\sqrt{2\lambda}} \;\leq\; \frac{\log L_b(\lambda)}{\sqrt{2\lambda}} \;\leq\; \frac{\log 2 + bM}{\sqrt{2\lambda}} - b.$$

Both sides converge to $-b$ as $\lambda \to \infty$, and hence, we have $-\log L_b(\lambda)/\sqrt{2\lambda} \to b$.

Step 2: From the empirical distribution to the population distribution. Note that $X_{n,i} := e^{-\lambda_n t_i} \in [0,1]$ are i.i.d. with mean $L_b(\lambda_n)$. We also denote the variance of $X_{n,i}$ by $\mathrm{Var}(X_{n,i})$. We next make the following claim.

Lemma 1. *Suppose the conditions in* (4) *hold. Then,* $\lim_{n\to\infty} n\,\mathrm{Var}(X_{n,i}) = \infty$.

Proof. First, by the law of total variance, it suffices to show this condition on any fixed drift $V = v$, given that

$$\mathrm{Var}(X_{n,i}) \geq \mathbb{E}_V\left[\mathrm{Var}(X_{n,i} \mid V = v)\right].$$

Next, note that we have

$$\begin{aligned}\mathrm{Var}(X_{n,i} \mid V = v) &= \mathbb{E}[e^{-2\lambda_n T} \mid V = v] - (\mathbb{E}[e^{-\lambda_n T} \mid V = v])^2 \\ &\geq \exp(-b\sqrt{M^2 + 4\lambda_n}) - 4\exp(2bM - 2b\sqrt{2\lambda_n}), \end{aligned} \tag{23}$$

where the last inequality follows from the same argument that we used to derive (21) and (22). Now, looking at (23), we note that, as $\lambda_n \to \infty$, ignoring the constants, the first term would decay as $\exp(-2b\sqrt{\lambda_n})$, while the second term would decay as $\exp(-2b\sqrt{2\lambda_n})$. Therefore, there exists n^* such that, for any $n \geq n^*$, we have

$$\exp(-b\sqrt{M^2 + 4\lambda_n}) - 4\exp(2bM - 2b\sqrt{2\lambda_n}) \geq \zeta\exp(-2b\sqrt{\lambda_n}), \tag{24}$$

for some constant $\zeta > 0$. As a result, for any $n \geq n^*$, we have

$$n\,\mathrm{Var}(X_{n,i}) \geq \zeta n\exp(-2b\sqrt{\lambda_n}). \tag{25}$$

The condition $\sqrt{\lambda_n} = o(\log n)$ completes the proof. □

Now, let us define $\tilde{X}_{n,i} = X_{n,i} - L_b(\lambda_n)$. We also define $s_n^2 := n\,\mathrm{Var}(\tilde{X}_{n,i}) = n\,\mathrm{Var}(X_{n,i})$ and $S_n := \sum_{i=1}^n \tilde{X}_{n,i}$. Note that, $X_{n,i} \leq 1$, and hence, $|\tilde{X}_{n,i}| \leq 1$.

The Lindeberg-Feller Central Limit Theorem [7][Theorem 27.2] states that, if the Lindeberg condition, given as

$$\lim_{n\to\infty} \frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}\left[\tilde{X}_{n,i}^2; |\tilde{X}_{n,i}| \geq \varepsilon s_n\right] = 0 \quad \text{for all } \varepsilon > 0. \tag{26}$$

holds, then, S_n/s_n converges to $\mathcal{N}(0,1)$ in distribution.

Notice that the Lindeberg condition (26) indeed holds, as $\tilde{X}_{n,i}$ is bounded and $s_n \to \infty$ by the above lemma. Therefore, we obtain that

$$\frac{S_n}{s_n} = \sqrt{n}\cdot\frac{\widehat{L}_n(\lambda_n) - L_b(\lambda_n)}{\sqrt{\mathrm{Var}(X_{n,i})}} \tag{27}$$

converges to $\mathcal{N}(0,1)$ in distribution. This implies that, for any $\tilde{M}$, we have

$$\lim_{n\to\infty} \Pr\left(\left|\sqrt{n}\cdot\frac{\widehat{L}_n(\lambda_n) - L_b(\lambda_n)}{\sqrt{\mathrm{Var}(X_{n,i})}}\right| \geq \tilde{M}\right) = 2(1 - \Phi(\tilde{M})), \tag{28}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Consequently, we have

$$\lim_{n\to\infty} \Pr\left(\left|\frac{\widehat{L}_n(\lambda_n)}{L_b(\lambda_n)} - 1\right| \geq \frac{\tilde{M}\sqrt{\mathrm{Var}(X_{n,i})}}{\sqrt{n}L_b(\lambda_n)}\right) = 2(1 - \Phi(\tilde{M})). \tag{29}$$

Note that $\mathrm{Var}(X_{n,i}) \leq L_b(\lambda_n)$, because

$$\mathrm{Var}(X_{n,i}) \leq \mathbb{E}[X_{n,i}^2] \leq \mathbb{E}[X_{n,i}] = L_b(\lambda_n), \tag{30}$$

and thus, we deduce from (29) that

$$\limsup_{n\to\infty} \Pr\left(\left|\frac{\widehat{L}_n(\lambda_n)}{L_b(\lambda_n)} - 1\right| \geq \frac{\tilde{M}}{\sqrt{nL_b(\lambda_n)}}\right) \leq 2(1 - \Phi(\tilde{M})), \tag{31}$$

for any $\tilde{M}$. Notice that, by (30), we have

$$\sqrt{nL_b(\lambda_n)} \geq \sqrt{n \operatorname{Var}(X_{n,i})}, \tag{32}$$

and therefore, by Lemma 1, $\sqrt{nL_b(\lambda_n)} \to \infty$ as n grows. Using this fact, we claim (31) implies that $\widehat{L}_n(\lambda_n)/L_b(\lambda_n) \xrightarrow{P} 1$. We can prove this by contradiction. Suppose this is not the case. Then, there exists positive ε and δ such that

$$\limsup_{n\to\infty} \Pr\left(\left|\frac{\widehat{L}_n(\lambda_n)}{L_b(\lambda_n)} - 1\right| \geq \varepsilon\right) \geq \delta. \tag{33}$$

However, setting $\tilde{M}$ in (31) such that $2(1-\Phi(\tilde{M})) < \delta$ leads to a contradiction as $\tilde{M}/\sqrt{nL_b(\lambda_n)}$ would fall below ε for large enough n. This completes the proof of our claim.

Having $\widehat{L}_n(\lambda_n)/L_b(\lambda_n) \xrightarrow{P} 1$ means $\log \widehat{L}_n(\lambda_n) - \log L_b(\lambda_n) \xrightarrow{P} 0$. Dividing by $\sqrt{2\lambda_n} \to \infty$ and combining with Step 1 gives $\widehat{b}_n = -\log \widehat{L}_n(\lambda_n)/\sqrt{2\lambda_n} \xrightarrow{P} b$.

A.2.1 Alternative proof of the second step

We next present an alternative proof of the second step. The original version of the proof above (for the second step) had a minor issue, which has now been fixed. This alternative proof was partially proposed, as a fix, by Google's Paper Assistant Tool, which NeurIPS made available to the authors. Although we fixed our proof in a different way from what the tool had suggested, we include this alternative proof for interested readers. We also present the proof in a self-contained manner and repeat the notation used in our proof of the second step for completeness.

Abbreviate

$$L_n := L_b(\lambda_n)$$

and define

$$X_{n,i} := e^{-\lambda_n T_i} \in [0,1].$$

The variables $X_{1,n}, \ldots, X_{n,n}$ are i.i.d. with mean L_n, and

$$\widehat{L}_n(\lambda_n) = \frac{1}{n}\sum_{i=1}^{n} X_{n,i}.$$

We first record that

$$\operatorname{Var}(X_{n,i}) \leq \mathbb{E}[X_{n,i}^2] \leq \mathbb{E}[X_{n,i}] = L_n,$$

where the second inequality uses $0 \leq X_{n,i} \leq 1$. Hence, for any $\varepsilon > 0$, Chebyshev's inequality gives

$$\begin{aligned}\mathbb{P}_b\left(\left|\frac{\widehat{L}_n(\lambda_n)}{L_b(\lambda_n)} - 1\right| > \varepsilon\right) &= \mathbb{P}_b\left(\left|\widehat{L}_n(\lambda_n) - L_n\right| > \varepsilon L_n\right) \\ &\leq \frac{\operatorname{Var}(X_{n,i})}{n\varepsilon^2 L_n^2} \leq \frac{1}{n\varepsilon^2 L_n}.\end{aligned}$$

It remains to show that $nL_n \to \infty$. By the lower bound from Step 1,

$$L_n = L_b(\lambda_n) \geq \exp\left(-b\sqrt{M^2 + 2\lambda_n}\right).$$

Therefore,

$$nL_n \geq \exp\left(\log n - b\sqrt{M^2 + 2\lambda_n}\right) \longrightarrow \infty,$$

because $\sqrt{\lambda_n} = o(\log n)$. Consequently,

$$\frac{\widehat{L}_n(\lambda_n)}{L_b(\lambda_n)} \xrightarrow{\mathbb{P}_b} 1.$$

Since $\widehat{L}_n(\lambda_n) > 0$ and $L_b(\lambda_n) > 0$, the continuous mapping theorem yields

$$\log \widehat{L}_n(\lambda_n) - \log L_b(\lambda_n) = \log\left(\frac{\widehat{L}_n(\lambda_n)}{L_b(\lambda_n)}\right) \xrightarrow{\mathbb{P}_b} 0.$$

Combining this with the population tail limit from Step 1,

$$-\frac{\log L_b(\lambda_n)}{\sqrt{2\lambda_n}} \longrightarrow b,$$

we obtain

$$\widehat{b}_n = -\frac{\log \widehat{L}_n(\lambda_n)}{\sqrt{2\lambda_n}} = -\frac{\log L_b(\lambda_n)}{\sqrt{2\lambda_n}} - \frac{\log\left(\widehat{L}_n(\lambda_n)/L_b(\lambda_n)\right)}{\sqrt{2\lambda_n}} \xrightarrow{\mathbb{P}_b} b.$$

This completes the proof of Theorem 3.2. □

A.3 Proof of Theorem 3

We establish the second-order expansion of $\log L_b(\lambda)$ and then transfer it to the empirical transform.

Step 1: Second-order expansion of $\log L_b(\lambda)$. Set $s := \sqrt{2\lambda}$ and, for $v \in [-M, M]$, define $R_s(v) := \sqrt{s^2 + v^2} - s = v^2/(\sqrt{s^2 + v^2} + s)$.

For $x \geq 0$,

$$\frac{1}{\cosh(x)} = 2e^{-x}(1 + \eta(x)), \tag{34}$$

with $|\eta(x)| \leq e^{-2x}$. Applying this with $x = b\sqrt{s^2 + v^2} = b(s + R_s(v))$, and using $\sqrt{s^2 + v^2} \geq s$, we obtain

$$\frac{1}{\cosh(b\sqrt{s^2 + v^2})} \;=\; 2e^{-bs}e^{-bR_s(v)}(1 + \eta), \qquad |\eta| \leq e^{-2bs}.$$

Multiplying by $\cosh(bv)$ and taking expectation over $V \sim F^*$ yields

$$L_b(\lambda) \;=\; 2e^{-bs}\,\mathbb{E}[\cosh(bV)e^{-bR_s(V)}] + O(e^{-3bs}).$$

Note that, there exist constants $K_1 = K_1(M)$, $K_2 = K_2(M)$ such that, uniformly on $|v| \leq M$ as $s \to \infty$,

$$\sup_{|v|\leq M} \left|R_s(v) - \frac{v^2}{2s}\right| \leq \frac{K_1}{s^3}, \qquad \sup_{|v|\leq M} |R_s(v)| \leq \frac{K_2}{s}.$$

Therefore, the following expansion holds uniformly on $|V| \leq M$

$$e^{-bR_s(V)} = 1 - bR_s(V) + O(R_s(V)^2) = 1 - \frac{bV^2}{2s} + O(s^{-2}), \tag{35}$$

which implies

$$\mathbb{E}[\cosh(bV)e^{-bR_s(V)}] \;=\; C_0 - \frac{bC_2}{2s} + O(s^{-2}), \tag{36}$$

with $C_0 = \mathbb{E}[\cosh(bV)] \geq 1$ and $C_2 = \mathbb{E}[V^2\cosh(bV)]$. Absorbing the $O(e^{-3bs})$ remainder into $O(s^{-2})e^{-bs}$ gives

$$L_b(\lambda) \;=\; 2C_0e^{-bs}\big(1 - \frac{bC_2}{2C_0s} + O(s^{-2})\big).$$

Taking logarithms and using $\log(1 + u) = u + O(u^2)$, we obtain

$$\log L_b(\lambda) \;=\; -bs + \log(2C_0) - \frac{bC_2}{2C_0s} + O(s^{-2}) \;=\; -b\sqrt{2\lambda} + \log(2C_0) - \frac{bC_2}{2C_0\sqrt{2\lambda}} + O(\lambda^{-1}), \tag{37}$$

which proves (7).

Step 2: Cancellation at two scales. Applying (7) at λ and 4λ (so $\sqrt{2(4\lambda)} = 2s$),

$$\log L_b(\lambda) - \log L_b(4\lambda) \;=\; bs - \frac{bC_2}{4C_0 s} + O(s^{-2}),$$

and dividing by $\sqrt{2\lambda} = s$,

$$\frac{\log L_b(\lambda) - \log L_b(4\lambda)}{\sqrt{2\lambda}} \;=\; b - \frac{bC_2}{8C_0\lambda} + O(\lambda^{-3/2}).$$

For comparison, $-\log L_b(\lambda)/\sqrt{2\lambda} \;=\; b - \log(2C_0)/\sqrt{2\lambda} + O(\lambda^{-1})$, so the one-scale bias is $O(\lambda^{-1/2})$ and the Richardson-corrected bias is $O(\lambda^{-1})$.

Step 3: The empirical estimator. Note that the sequence $\{4\lambda_n\}$ also satisfies (4), and hence, by Theorem 2,

$$-\frac{\log \widehat{L}_n(\lambda_n)}{\sqrt{2\lambda_n}} \xrightarrow{P} b, \quad -\frac{\log \widehat{L}_n(4\lambda_n)}{2\sqrt{2\lambda_n}} \xrightarrow{P} b \tag{38}$$

This immediately implies

$$\frac{\log \widehat{L}_n(\lambda_n) - \log \widehat{L}_n(4\lambda_n)}{\sqrt{2\lambda_n}} \xrightarrow{P} b \tag{39}$$

which completes the proof. □

A.4 Proof of Theorem 4

Fix a realized boundary $b > 0$ and write $\mathbb{P}_b, \mathbb{E}_b$ for the conditional probability and expectation. The proof relies on a uniform envelope for the family $\{w_\beta\}$, stated next, and whose proof we give at the end of this subsection.

Lemma 2 (Uniform envelope for the weight family)**.** *Fix a compact interval $K = [\underline{b}, \bar{b}] \subset (0, \infty)$. Then:*

1. *For every $\tau > 0$, $\beta \mapsto w_\beta(\tau)$ is continuous on K.*
2. *There exist constants $0 < \tau_0 \le 1$, $\tau_1 \ge 1$, $C_K < \infty$, and $c_K > 0$, all depending only on K such that, for all $\beta \in K$, we have*

$$\begin{aligned} &|w_\beta(\tau)| \le C_K(1 + \tau^{-1}) && \text{for } 0 < \tau \le \tau_0, \\ &|w_\beta(\tau)| \le C_K && \text{for } \tau_0 \le \tau \le \tau_1, \\ &|w_\beta(\tau)| \le C_K e^{-c_K \tau} && \text{for } \tau \ge \tau_1. \end{aligned}$$

Consequently, $H_K(\tau) := C_K(1 + \tau^{-1})\mathbf{1}_{(0,\tau_0]}(\tau) + C_K \mathbf{1}_{(\tau_0,\tau_1]}(\tau) + C_K e^{-c_K\tau}\mathbf{1}_{[\tau_1,\infty)}(\tau)$ satisfies $\sup_{\beta\in K} |w_\beta(\tau)| \le H_K(\tau)$, and $\mathbb{E}_b[H_K(T)] < \infty$.

Main argument. Decompose

$$\widehat{\mu}_n^{\mathrm{PI}} - \mu = \underbrace{\left(\frac{1}{n}\sum_{i=1}^n z_i w_b(t_i) - \mu\right)}_{A_n} + \underbrace{\frac{1}{n}\sum_{i=1}^n z_i \big(w_{\widetilde{B}_n}(t_i) - w_b(t_i)\big)}_{R_n}.$$

By Theorem 1, we know that $\mathbb{E}_b[z_i w_b(t_i)] = \mu$. Furthermore, Lemma 2 gives $\mathbb{E}_b[|z_i w_b(t_i)|] \le \mathbb{E}_b[H_K(t_i)] < \infty$ for any compact $K \ni b$. Hence, by the weak law of large numbers, we have $A_n \xrightarrow{\mathbb{P}_b} 0$.

For R_n, pick the compact set $K = [b/2, 3b/2]$ and, for $0 < \delta \le b/2$, define

$$r_\delta(\tau) := \sup_{|\beta - b| \le \delta} |w_\beta(\tau) - w_b(\tau)|. \tag{40}$$

Lemma 2 implies that (i) $r_\delta(\tau) \downarrow 0$ pointwise as $\delta \downarrow 0$, and (ii) $r_\delta(\tau) \le 2H_K(\tau)$ where H_K is integrable. Hence, by dominated convergence:

$$\lim_{\delta \to 0} \mathbb{E}_b[r_\delta(T)] = 0. \tag{41}$$

Fix $\varepsilon > 0$ and pick $\delta \in (0, b/2]$ with $\mathbb{E}_b[r_\delta(T)] < \varepsilon/2$. On the event $\{|\widetilde{B}_n - b| \le \delta\}$, we have

$$|R_n| \le \frac{1}{n} \sum_i r_\delta(\tau_i).$$

Therefore,

$$\mathbb{P}_b(|R_n| > \varepsilon) \;\le\; \mathbb{P}_b(|\widetilde{B}_n - b| > \delta) + \mathbb{P}_b\left(\frac{1}{n} \sum_i r_\delta(\tau_i) > \varepsilon \right).$$

The first term vanishes by Theorem 3 and the second by the weak law of large numbers (the sample mean tends to $\mathbb{E}_b[r_\delta(\tau_i)] < \varepsilon/2$). So $R_n \xrightarrow{\mathbb{P}_b} 0$, and $\widehat{\mu}_n^{\mathrm{PI}} - \mu = A_n + R_n \xrightarrow{\mathbb{P}_b} 0$. This completes the proof. □

A.4.1 Proof of Lemma 2.

Continuity. Recall that $w_\beta(t)$ is equal to $N_\beta(t)/D_\beta(t)$, with

$$N_\beta(t) := \sum_{k \ge 0} \exp\left(\frac{-c_k(\beta)^2}{2t} \right) \left(\frac{c_k(\beta)^2}{t} - 1 \right)$$

and

$$D_\beta(t) := \sum_{k \ge 0} (-1)^k c_k(\beta) \exp\left(\frac{-c_k(\beta)^2}{2t} \right).$$

For a fixed $t > 0$, both of these series converge absolutely and uniformly on K because, for $m \in \{1, 2\}$,

$$\sum_{k \ge 0} (2k+1)^m e^{-(2k+1)^2 \underline{b}^2/(2t)} < \infty.$$

Using the known identity $D_\beta(t) = \sqrt{\pi/2}\, t^{3/2} f_{0,\beta}(t)$ with $f_{0,\beta}$ being the first-exit density for boundary β and zero drift (see Eq. (17)), we have $D_\beta(t) > 0$ for every $t > 0$, so $w_\beta(t) = N_\beta(t)/D_\beta(t)$ is continuous in β on K.

Small-t bound. Factor

$$D_\beta(t) = \beta e^{-\beta^2/(2t)} \big[1 + \sum_{k \ge 1} (-1)^k (2k+1) e^{-((2k+1)^2 - 1)\beta^2/(2t)} \big].$$

Since $((2k+1)^2 - 1)\beta^2/(2t) \ge 4\underline{b}^2/t$ for $k \ge 1$, we can choose $\tau_0 > 0$ so that

$$\sup_{\beta \in K} \sum_{k \ge 1} (2k+1) e^{-((2k+1)^2 - 1)\beta^2/(2t)} \le 1/2$$

on $(0, \tau_0]$. Then $D_\beta(t) \ge \frac{1}{2}\beta e^{-\beta^2/(2t)} \ge \frac{1}{2}\underline{b} e^{-\beta^2/(2t)}$.

Now control the numerator. Factor out $e^{-\beta^2/(2t)}$:

$$N_\beta(t) = e^{-\beta^2/(2t)} \left[\frac{\beta^2}{t} - 1 + \sum_{k=1}^{\infty} e^{-((2k+1)^2 - 1)\beta^2/(2t)} \left(\frac{(2k+1)^2 \beta^2}{t} - 1 \right) \right].$$

The first term satisfies $\left| \frac{\beta^2}{t} - 1 \right| \le C_K(1 + t^{-1})$. The tail satisfies

$$\sum_{k=1}^{\infty} e^{-((2k+1)^2 - 1)\beta^2/(2t)} \left| \frac{(2k+1)^2 \beta^2}{t} - 1 \right| \le C_K(1 + t^{-1}),$$

because the sums

$$\sum_{k=1}^{\infty} (2k+1)^2 e^{-((2k+1)^2 - 1) b^2/(2t)}$$

are uniformly bounded for $0 < t \le \tau_0$.

Therefore, $|N_\beta(t)| \le C_K e^{-\beta^2/(2t)}(1 + t^{-1})$, so $|w_\beta(t)| \le C_K(1 + t^{-1})$.

Middle range. Continuity of $(\beta, \tau) \mapsto w_\beta(\tau)$ on the compact $K \times [\tau_0, \tau_1]$ yields $\sup |w_\beta(\tau)| < \infty$.

Large-t bound. Note that

$$w_\beta(t) = \frac{N_\beta(t)}{D_\beta(t)} = \frac{\sqrt{\frac{2}{\pi}} t^{-3/2} N_\beta(t)}{f_{0,\beta}(t)}.$$

We next make two claims:

Lemma 3. *Recall that $f_{0,\beta}(t)$ is given by*

$$f_{0,\beta}(t) = \sqrt{\frac{2}{\pi}}\, t^{-3/2} \sum_{k=0}^{\infty} (-1)^k c_k(\beta) \exp\left(-\frac{c_k(\beta)^2}{2t}\right). \tag{42}$$

This function also admits the following characterization:

$$f_{0,\beta}(t) = \frac{\pi}{2\beta^2} \sum_{m \geq 0} (-1)^m (2m+1) \exp\left(-\frac{(2m+1)^2 \pi^2 t}{8\beta^2}\right). \tag{43}$$

Lemma 4. *Define $q_\beta(t)$ by*

$$q_\beta(t) := \sqrt{\frac{2}{\pi}} t^{-3/2} N_\beta(t) = \sqrt{\frac{2}{\pi}} t^{-3/2} \sum_{k \geq 0} \exp\left(\frac{-c_k(\beta)^2}{2t}\right) \left(\frac{c_k(\beta)^2}{t} - 1\right). \tag{44}$$

This function also admits the following characterization:

$$q_\beta(t) = \frac{\pi^2}{\beta^3} \sum_{m \geq 1} (-1)^{m+1} m^2 \exp\left(-\frac{m^2 \pi^2 t}{2\beta^2}\right) \tag{45}$$

We defer the proof of these two claims to the end of this subsection. First let us complete the proof of Lemma 2, assuming these two results hold.

Note that, for $t \geq \tau_1$ large enough, the $m = 0$ term of $f_{0,\beta}$ and the $m = 1$ term of q_β dominate, giving $f_{0,\beta}(t) \geq \frac{\pi}{4\beta^2} e^{-\pi^2 t/(8\beta^2)}$ and $|q_\beta(t)| \leq C_K e^{-\pi^2 t/(2\beta^2)}$. Hence

$$|w_\beta(t)| \;=\; \frac{|q_\beta(t)|}{f_{0,\beta}(t)} \;\leq\; C_K \exp\!\big(-\frac{3\pi^2 t}{8\beta^2}\big) \;\leq\; C_K e^{-c_K t}, \qquad c_K := \frac{3\pi^2}{8\bar{b}^2}.$$

Combining the three regions gives the envelope.

Integrability of H_K. Let g_b be the marginal density of T under $\mathbb{P}_b$. Summing (13) over $z \in \{\pm 1\}$ gives $g_b(t) = \mathbb{E}[\cosh(bV) e^{-V^2 t/2}] f_{0,b}(t) \leq \cosh(bM) f_{0,b}(t)$, and for small t, the formula for $f_{0,b}$ yields $f_{0,b}(t) \leq C t^{-3/2} e^{-b^2/(2t)}$. Hence

$$\int_0^{\tau_0} (1 + \tau^{-1}) g_b(\tau)\, d\tau \;\leq\; C \int_0^{\tau_0} (\tau^{-3/2} + \tau^{-5/2}) e^{-b^2/(2\tau)}\, d\tau \;<\; \infty,$$

because $e^{-b^2/(2\tau)}$ decays faster than any inverse power as $\tau \downarrow 0$. On $[\tau_0, \infty)$, H_K is bounded, so its integral against g_b is finite. Thus $\mathbb{E}_b[H_K(t_i)] < \infty$. □

A.4.2 Proof of Lemma 3

First note that, the Laplace transform of each term of (42) is given by

$$\mathcal{L}\left[\sqrt{\frac{2}{\pi}} t^{-3/2}\; a\; e^{-a^2/(2t)}\right](\lambda) = 2e^{-a\sqrt{2\lambda}}, \qquad a > 0,\ \lambda > 0. \tag{46}$$

This implies that, the Laplace transform of $f_{0,\beta}(t)$ in (42) is given by

$$\mathcal{L}[f](\lambda) = \sum_{k=0}^{\infty} (-1)^k\; 2e^{-(2k+1)\beta\sqrt{2\lambda}} = \operatorname{sech}(\beta\sqrt{2\lambda}). \tag{47}$$

Now, the Mittag–Leffler expansion for the $\mathrm{sech}(\cdot)$ gives us

$$\mathcal{L}[f](\lambda) = \pi \sum_{m=0}^{\infty} (-1)^m \frac{2m+1}{2\beta^2\lambda + \frac{\pi^2}{4}(2m+1)^2} = \frac{\pi}{2\beta^2} \sum_{m=0}^{\infty} (-1)^m (2m+1) \frac{1}{\lambda + \frac{(2m+1)^2\pi^2}{8\beta^2}}. \tag{48}$$

Taking an inverse Laplace transform and using the basic fact that

$$\mathcal{L}^{-1}\left[\frac{1}{\lambda+\alpha}\right](t) = e^{-\alpha t}, \qquad \alpha > 0,$$

gives us the desired expression (43). □

A.4.3 Proof of Lemma 4

Note that $N_\beta(t)$ can be cast as

$$N_\beta(t) = -\partial_\beta\big(\beta S_\beta(t)\big) \tag{49}$$

with

$$S_\beta(t) = \sum_{k=0}^{\infty} e^{-((2k+1)^2\beta^2)/(2t)}. \tag{50}$$

It is worth noting that we are allowed to use termwise differentiation given $\beta \geq \underline{b} > 0$. We next compute $S_\beta(t)$ by the Jacobi theta transformation, equivalently Poisson summation. Since

$$2S_\beta(t) = \sum_{n\in\mathbb{Z}} \exp\left(-\frac{(2n+1)^2\beta^2}{2t}\right) = \sum_{n\in\mathbb{Z}} \exp\left(-\frac{2\beta^2}{t}\left(n+\frac{1}{2}\right)^2\right), \tag{51}$$

we apply the shifted Poisson summation formula

$$\sum_{n\in\mathbb{Z}} f(n+\alpha) = \sum_{m\in\mathbb{Z}} e^{2\pi i m\alpha}\, \widehat{f}(m), \tag{52}$$

with

$$f(x) := \exp\left(-\frac{2\beta^2}{t}x^2\right), \qquad \alpha := \frac{1}{2}. \tag{53}$$

Using the Fourier transform convention

$$\widehat{f}(\xi) := \int_{\mathbb{R}} f(x) e^{-2\pi i x\xi}\, dx, \tag{54}$$

the Gaussian transform is

$$\widehat{f}(\xi) = \sqrt{\frac{\pi t}{2\beta^2}}\, \exp\left(-\frac{\pi^2 t}{2\beta^2}\xi^2\right). \tag{55}$$

Substituting (53) and (55) into (52), and then using (51), gives

$$\begin{aligned} 2S_\beta(t) &= \sqrt{\frac{\pi t}{2\beta^2}} \sum_{m\in\mathbb{Z}} e^{2\pi i m/2} \exp\left(-\frac{\pi^2 t}{2\beta^2}m^2\right) \\ &= \sqrt{\frac{\pi t}{2\beta^2}} \sum_{m\in\mathbb{Z}} (-1)^m \exp\left(-\frac{\pi^2 t}{2\beta^2}m^2\right). \end{aligned} \tag{56}$$

Since the summand is even in m,

$$\sum_{m\in\mathbb{Z}} (-1)^m e^{-am^2} = 1 + 2\sum_{m=1}^{\infty} (-1)^m e^{-am^2}, \qquad a > 0. \tag{57}$$

Applying (57) with

$$a = \frac{\pi^2 t}{2\beta^2}, \tag{58}$$

we find

$$S_\beta(t) = \frac{1}{2}\sqrt{\frac{\pi t}{2\beta^2}}\left(1 + 2\sum_{m=1}^{\infty}(-1)^m \exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right)\right)$$
$$= \frac{\sqrt{\pi t/2}}{2\beta}\left(1 + 2\sum_{m=1}^{\infty}(-1)^m \exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right)\right), \tag{59}$$

because $\beta > 0$.

Multiplying (59) by β yields

$$\beta S_\beta(t) = \frac{\sqrt{\pi t/2}}{2}\left(1 + 2\sum_{m=1}^{\infty}(-1)^m \exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right)\right). \tag{60}$$

We now differentiate (60) with respect to β. For each $m \geq 1$, define

$$A_m := \frac{m^2\pi^2 t}{2}. \tag{61}$$

Then

$$\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right) = e^{-A_m/\beta^2}, \tag{62}$$

and therefore

$$\frac{d}{d\beta}e^{-A_m/\beta^2} = e^{-A_m/\beta^2}\,\frac{d}{d\beta}\left(-A_m\beta^{-2}\right)$$
$$= e^{-A_m/\beta^2}\,\left(2A_m\beta^{-3}\right)$$
$$= \frac{m^2\pi^2 t}{\beta^3}\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right). \tag{63}$$

Differentiating (60) term-by-term, we obtain

$$\frac{\partial}{\partial\beta}(\beta S_\beta(t)) = \frac{\sqrt{\pi t/2}}{2}\cdot 2\sum_{m=1}^{\infty}(-1)^m\frac{m^2\pi^2 t}{\beta^3}\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right)$$
$$= \sqrt{\frac{\pi t}{2}}\,\frac{\pi^2 t}{\beta^3}\sum_{m=1}^{\infty}(-1)^m m^2\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right). \tag{64}$$

Using (49), we conclude that

$$N_\beta(t) = \sqrt{\frac{\pi t}{2}}\,\frac{\pi^2 t}{\beta^3}\sum_{m=1}^{\infty}(-1)^{m+1}m^2\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right). \tag{65}$$

Finally, multiply (65) by $\sqrt{2/\pi}\,t^{-3/2}$:

$$\sqrt{\frac{2}{\pi}}\,t^{-3/2}N_\beta(t) = \sqrt{\frac{2}{\pi}}\,t^{-3/2}\cdot\sqrt{\frac{\pi t}{2}}\cdot\frac{\pi^2 t}{\beta^3}\sum_{m=1}^{\infty}(-1)^{m+1}m^2\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right)$$
$$= \left(\sqrt{\frac{2}{\pi}}\sqrt{\frac{\pi t}{2}}t^{-3/2}\pi^2 t\right)\frac{1}{\beta^3}\sum_{m=1}^{\infty}(-1)^{m+1}m^2\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right)$$
$$= \left(t^{1/2}t^{-3/2}\pi^2 t\right)\frac{1}{\beta^3}\sum_{m=1}^{\infty}(-1)^{m+1}m^2\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right)$$
$$= \frac{\pi^2}{\beta^3}\sum_{m=1}^{\infty}(-1)^{m+1}m^2\exp\left(-\frac{m^2\pi^2 t}{2\beta^2}\right). \tag{66}$$

This is exactly the claimed identity (45). □

A.5 Proof of Theorem 5

Throughout this proof, fix $b > 0$ and define $\mathbb{P}_b$ and $\mathbb{E}_b$ as in Appendix A.4.

Step 1: Boundary consistency. Because $\|\psi_i\| \le L$ and $\|\theta_i\| \le R$, $|v_i| = |\psi_i^\top \theta_i| \le LR$ almost surely, so the scalar drifts $\{v_i\}$ satisfy the bounded-drift hypothesis of Theorem 3. Hence $\widetilde{B}_n \xrightarrow{\mathbb{P}_b} b$.

Step 2: Covariance matrix. Since $\mathbb{E}\|\psi_i\psi_i^\top\|_{\mathrm{op}} \le L^2 < \infty$, the weak law of large numbers gives $\widehat{Q}_n \xrightarrow{\mathbb{P}_b} Q$.

Step 3: Moment convergence. Decompose $\widehat{m}_n = S_n + R_n$ with $S_n := \frac{1}{n}\sum_i \psi_i z_i w_b(t_i)$ and $R_n := \frac{1}{n}\sum_i \psi_i z_i (w_{\widetilde{B}_n}(t_i) - w_b(t_i))$.

Analysis of S_n. Conditional on (ψ_i, θ_i), the drift is the fixed scalar $v_i = \psi_i^\top \theta_i$, so Theorem 1 gives $\mathbb{E}_b[z_i w_b(t_i) \mid \psi_i, \theta_i] = \psi_i^\top \theta_i$. Multiplying by ψ_i and applying the exogeneity condition $\mathbb{E}_b[\theta_i \mid \psi_i] = \theta^\star$,

$$\mathbb{E}_b[\psi_i z_i w_b(t_i)] \;=\; \mathbb{E}_b[\psi_i \psi_i^\top \theta_i] \;=\; \mathbb{E}_b[\psi_i \psi_i^\top \mathbb{E}_b[\theta_i \mid \psi_i]] \;=\; Q\theta^\star.$$

Also $\|\psi_i z_i w_b(t_i)\| \le L|w_b(t_i)| \le L H_K(t_i)$ for $K = [b/2, 3b/2]$, which is integrable by Lemma 2. Hence $S_n \xrightarrow{\mathbb{P}_b} Q\theta^\star$ by the weak law of large numbers.

Analysis of R_n. For $\delta \in (0, b/2]$, define r_δ as in the scalar proof. On $\{|\widetilde{B}_n - b| \le \delta\}$, $\|R_n\| \le L\frac{1}{n}\sum_i r_\delta(t_i)$. Choose δ with $L\,\mathbb{E}_b[r_\delta(t_i)] < \varepsilon/2$. Then

$$\mathbb{P}_b(\|R_n\| > \varepsilon) \;\le\; \mathbb{P}_b(|\widetilde{B}_n - b| > \delta) + \mathbb{P}_b\Big(\frac{L}{n}\sum_i r_\delta(t_i) > \varepsilon\Big),$$

and both terms vanish by Theorem 3 and the weak law of large numbers respectively. So $R_n \xrightarrow{\mathbb{P}_b} 0$.

Step 4: Continuous mapping. Combining $\widehat{Q}_n \xrightarrow{\mathbb{P}_b} Q$ with $\widehat{m}_n = S_n + R_n \xrightarrow{\mathbb{P}_b} Q\theta^\star$ and applying the continuous mapping theorem to $(\widehat{Q}_n, \widehat{m}_n) \mapsto \widehat{Q}_n^{-1}\widehat{m}_n$ on the open set $\{Q : \det Q > 0\}$,

$$\widehat{\theta}_n \;=\; \widehat{Q}_n^{-1}\widehat{m}_n \;\xrightarrow{\mathbb{P}_b}\; Q^{-1}(Q\theta^\star) \;=\; \theta^\star.$$

For every fixed x, $\widehat{\theta}_n^\top \phi(x) \xrightarrow{\mathbb{P}_b} \theta^{\star\top}\phi(x)$ by the continuous-mapping theorem. □

B Additional Results and Discussion

In this section, we provide additional results and further discussion, mainly related to some of the assumptions made in the paper.

B.1 Non-decision time

In some applications, it is natural to assume that the decision-maker takes some time to parse the two alternatives before starting the evidence accumulation that is modeled by means of the DDM. In particular, the classical **extended** DDM [56] models such non-decision time as an additive term. More specifically, the observed response time, now denoted by S, is given by

$$S = T + A, \tag{67}$$

where T is the latent DDM first-passage time and $A \ge 0$ denotes the non-decision time. It is worth noting that, here, A is not assumed to be a function of the alternatives; rather, it may be either deterministic or stochastic.

In this subsection, we show how our model can be adjusted to handle a deterministic additive non-decision time. When A is allowed to be random or heterogeneous, additional complexity arises since, in that case, one must learn the boundary, the heterogeneous drift, and the random non-decision

response time jointly, while potentially having only one data point per labeler. One can use our results to derive bounds on the estimation error in that case (see Section B.2); we leave this direction for future work.

Before stating our results, it is useful to highlight that, in settings such as LLM responses, longer responses are typically harder to parse and compare. In this case, one can take a different approach than modeling the non-decision response time as an additive term and instead incorporate it into the drift variable. Our linear model in Section 4 already handles this scenario: it suffices to add one additional coordinate to ψ_i, whose corresponding coordinate in θ_i models this effect.

B.1.1 Extended DDM with non-decision time

Throughout this subsection, we assume $A = a > 0$ is deterministic. Let

$$L_T(\lambda) := \mathbb{E}\left[e^{-\lambda T}\right], \qquad L_S(\lambda) := \mathbb{E}\left[e^{-\lambda S}\right] = e^{-a\lambda} L_T(\lambda).$$

First, we argue why we need to adjust the estimator of the boundary, and why the two-scale estimator (5) fails. Note that, the population version of this estimator satisfies

$$\frac{\log L_S(\lambda) - \log L_S(4\lambda)}{\sqrt{2\lambda}} = \frac{3a}{\sqrt{2}}\sqrt{\lambda} + \frac{\log L_T(\lambda) - \log L_T(4\lambda)}{\sqrt{2\lambda}},$$

and even though we know the second term converges to b as $\lambda \to \infty$, the whole estimator would diverge to $+\infty$ at rate $\sqrt{\lambda}$. This observation motivates a three-scale estimator that cancels the linear non-decision-time term.

Proposition 1 (Three-scale boundary and non-decision-time estimators)**.** *Assume that $|V| \le M$ almost surely and that $\{\lambda_n\}_n$ satisfies* (4)*. Moreover, assume that $S = T + a$, where $a \ge 0$ is deterministic. Let*

$$\widehat{L}_n^S(\lambda) := \frac{1}{n}\sum_{i=1}^{n} e^{-\lambda S_i}, \qquad \widehat{F}_n(\lambda) := \log \widehat{L}_n^S(\lambda).$$

Define

$$\widehat{b}_n^{\mathrm{NDT}} := \frac{\frac{5}{2}\widehat{F}_n(\lambda_n) - 4\widehat{F}_n(4\lambda_n) + \frac{3}{2}\widehat{F}_n(9\lambda_n)}{\sqrt{2\lambda_n}},$$

and

$$\widehat{a}_n := \frac{-\frac{1}{2}\widehat{F}_n(\lambda_n) + \widehat{F}_n(4\lambda_n) - \frac{1}{2}\widehat{F}_n(9\lambda_n)}{\lambda_n}.$$

Then,

$$\widehat{b}_n^{\mathrm{NDT}} \xrightarrow{P} b, \qquad \widehat{a}_n \xrightarrow{P} a.$$

Proof. The proof follows the logic of the proof of Theorem 3 in Appendix A.3. Using the expansion of $L_b(\lambda)$ developed there, it is straightforward to verify that the population versions of $\widehat{b}_n^{\mathrm{NDT}}$ and $\widehat{a}_n$ converge to the corresponding parameters as λ_n grows. To obtain convergence in population, since $A = a$ is deterministic, we can again use the result developed in Theorem 2, as in the proof of Theorem 3. □

We conclude this subsection by deriving an analogue of Theorem 4 and designing a consistent estimator of μ while accounting for the non-decision time. The main idea is to remove the effect of the non-decision time by using the consistent estimator $\widehat{a}_n$ developed above. However, since $\widehat{a}_n$ is itself an estimator, it may even exceed some observed response times. In particular, the next result shows how to address this difficulty carefully.

Theorem 6. *Under the premise of Proposition 1, let $\varepsilon_n \downarrow 0$ be any deterministic sequence satisfying*

$$Pr(|\widehat{a}_n - a| > \varepsilon_n/2) \to 0. \tag{68}$$

Define

$$\widehat{\mu}_{n,\varepsilon}^{\mathrm{NDT}} := \frac{1}{n}\sum_{i=1}^{n} Z_i w_{\widehat{b}_n^{\mathrm{NDT}}}(S_i - \widehat{a}_n)\mathbf{1}\{S_i - \widehat{a}_n \ge \varepsilon_n\}. \tag{69}$$

Then, $\widehat{\mu}_{n,\varepsilon}^{\mathrm{NDT}} \xrightarrow{P} \mu = \mathbb{E}[V]$.

Proof. In this proof, we make use of the following lemma.

Lemma 5. *Let T be the first passage time of DDM, i.e., the response time under zero non-decision time. Then, we have*

$$\lim_{\varepsilon \downarrow 0} \mathbb{E}[|w_b(T)| \mathbf{1}\{T \le 2\varepsilon\}] = 0 \tag{70}$$

$$\lim_{\rho \downarrow 0} \limsup_{\varepsilon \downarrow 0} \mathbb{E}\left[\sup_{\substack{|\alpha - b| \le \rho \\ |u| \le \varepsilon \\ T+u \ge \varepsilon}} |w_\alpha(T+u) - w_b(T)| \right] = 0. \tag{71}$$

We first show how to prove Theorem 6 using this lemma, and then prove the lemma itself. Define $u_n := a - \widehat{a}_n$ as the error of the non-decision time estimator a at time n. Then, we have

$$S_i - \widehat{a}_n = T_i + a - \widehat{a}_n = T_i + u_n.$$

Let $\rho > 0$ be fixed and define the good event

$$E_{n,\rho} := \left\{ |\widehat{b}_n^{\mathrm{NDT}} - b| \le \rho, \quad |\widehat{a}_n - a| \le \varepsilon_n/2 \right\}. \tag{72}$$

By $\widehat{b}_n^{\mathrm{NDT}} \xrightarrow{P} b$ and (68),

$$\Pr(E_{n,\rho}) \to 1 \qquad \text{for every fixed } \rho > 0.$$

Decompose

$$\widehat{\mu}_{n,\varepsilon}^{\mathrm{NDT}} - \mu = A_n + B_n,$$

where

$$A_n := \frac{1}{n} \sum_{i=1}^n Z_i w_b(T_i) - \mu$$

and

$$B_n := \frac{1}{n} \sum_{i=1}^n Z_i \left[w_{\widehat{b}_n^{\mathrm{NDT}}}(T_i + u_n) \mathbf{1}\{T_i + u_n \ge \varepsilon_n\} - w_b(T_i) \right].$$

We already know $A_n \xrightarrow{P} 0$. Hence, it suffices to show $B_n \xrightarrow{P} 0$.

Notice that, on the event $E_{n,\rho}$, if $T_i + u_n < \varepsilon_n$, then $T_i < 3\varepsilon_n/2$. Hence, using $|Z_i| = 1$, on the event $E_{n,\rho}$, we have

$$|B_n| \le \frac{1}{n} \sum_{i=1}^n \sup_{\substack{|\alpha - b| \le \rho \\ |u| \le \varepsilon_n/2 \\ T_i + u \ge \varepsilon_n 0}} |w_\alpha(T_i + u) - w_b(T_i)| + \frac{1}{n} \sum_{i=1}^n |w_b(T_i)| \mathbf{1}\{T_i < 3\varepsilon_n/2\}.$$

As a result, for any ρ, we have

$$\begin{aligned}
\Pr(|B_n| > \delta) &\le \Pr(E_{n,\rho}) + \Pr\left(\frac{1}{n} \sum_{i=1}^n \sup_{\substack{|\alpha - b| \le \rho \\ |u| \le \varepsilon_n/2 \\ T_i + u \ge \varepsilon_n 0}} |w_\alpha(T_i + u) - w_b(T_i)| \ge \frac{\delta}{2} \right) \\
&\quad + \Pr\left(\frac{1}{n} \sum_{i=1}^n |w_b(T_i)| \mathbf{1}\{T_i < 3\varepsilon_n/2\} \ge \frac{\delta}{2} \right) \\
&\overset{\text{(Markov's inequality)}}{\le} \Pr(E_{n,\rho}) + \frac{2}{\delta} \mathbb{E}\left[\sup_{\substack{|\alpha - b| \le \rho \\ |u| \le \varepsilon_n/2 \\ T_i + u \ge \varepsilon_n 0}} |w_\alpha(T + u) - w_b(T)| \right]
\end{aligned}$$

$$+ \frac{2}{\delta}\mathbb{E}\left[|w_b(T)|\mathbf{1}\{T < 3\varepsilon_n/2\}\right]. \tag{73}$$

Now, we claim that this implies $\Pr(|B_n| > \delta)$ goes to zero as $n \to \infty$. Suppose not. Then there exist $\zeta > 0$ and an infinite subsequence of values of n such that $\Pr(|B_n| > \delta) > \zeta$.

Notice that, by Lemma 5, the last term in (73) goes to zero as $n \to \infty$. Also, again by Lemma 5, one can choose ρ small enough to make the second term arbitrarily small. Finally, for any fixed ρ, the first term goes to zero as $n \to \infty$. Taken together, these facts yield a contradiction and hence complete the proof of Theorem 6. It remains to prove Lemma 5.

Proof of Lemma 5 Let us define

$$\Delta_{\rho,\varepsilon}(t) := \sup_{\substack{|\alpha-b|\le\rho \\ |u|\le\varepsilon \\ t+u\ge\varepsilon}} |w_\alpha(t+u) - w_b(t)|.$$

Let H_K be the envelope from Lemma 2, so that

$$\sup_{\beta\in K} |w_\beta(t)| \le H_K(t), \qquad \mathbb{E}[H_K(T)] < \infty.$$

The limit (70) follows immediately from dominated convergence:

$$|w_b(T)|\mathbf{1}\{T \le 2\varepsilon\} \le H_K(T),$$

and

$$H_K(T)\mathbf{1}\{T \le 2\varepsilon\} \to 0 \qquad \text{almost surely as } \varepsilon \downarrow 0.$$

Since $H_K(T)$ is integrable, dominated convergence gives

$$\lim_{\varepsilon\downarrow 0} \mathbb{E}[|w_b(T)|\mathbf{1}\{T \le 2\varepsilon\}] = 0.$$

We now prove (71). Decompose

$$\mathbb{E}[\Delta_{\rho,\varepsilon}(T)] = \mathbb{E}[\Delta_{\rho,\varepsilon}(T)\mathbf{1}\{T > 2\varepsilon\}] + \mathbb{E}[\Delta_{\rho,\varepsilon}(T)\mathbf{1}\{T \le 2\varepsilon\}].$$

First consider the event $\{T > 2\varepsilon\}$. On this event, for every $|u| \le \varepsilon$,

$$T + u \ge T - \varepsilon > T/2.$$

Assume also $\varepsilon \le 1$. Then

$$T + u \le T + 1.$$

Define

$$\widetilde{H}_K(t) := H_K(t) + \sup_{s\in[t/2,t+1]} H_K(s).$$

By Lemma 2, $\widetilde{H}_K(T)$ is integrable. Indeed, for small t,

$$\sup_{s\in[t/2,t+1]} H_K(s) \le C(1 + t^{-1}),$$

and the small-time integrability follows from Lemma 2. For t bounded away from zero, H_K is bounded by the middle-region and large-time pieces of Lemma 2. Therefore, on $\{T > 2\varepsilon\}$,

$$\Delta_{\rho,\varepsilon}(T) \le \widetilde{H}_K(T).$$

For each fixed $t > 0$,

$$\lim_{\varepsilon\downarrow 0} \Delta_{\rho,\varepsilon}(t)\mathbf{1}\{t > 2\varepsilon\} = r_\rho(t),$$

where

$$r_\rho(t) := \sup_{|\alpha-b|\le\rho} |w_\alpha(t) - w_b(t)|.$$

The pointwise convergence follows from joint continuity of $(\alpha, t) \mapsto w_\alpha(t)$ on compact subsets of $K \times (0, \infty)$. Thus dominated convergence gives

$$\lim_{\varepsilon\downarrow 0} \mathbb{E}[\Delta_{\rho,\varepsilon}(T)\mathbf{1}\{T > 2\varepsilon\}] = \mathbb{E}[r_\rho(T)].$$

By Lemma 2,

$$r_\rho(T) \le 2H_K(T),$$

and

$$r_\rho(T) \to 0 \qquad \text{almost surely as } \rho \downarrow 0.$$

Another application of dominated convergence gives

$$\lim_{\rho \downarrow 0} \mathbb{E}[r_\rho(T)] = 0.$$

It remains to control the event $\{T \le 2\varepsilon\}$. On this event, if $|u| \le \varepsilon$ and $T + u \ge \varepsilon$, then

$$\varepsilon \le T + u \le 3\varepsilon.$$

For ε small enough that $3\varepsilon \le \tau_0$, Lemma 2 gives

$$|w_\alpha(T+u)| \le C_K(1+\varepsilon^{-1}) \qquad \text{for all } \alpha \in K.$$

Therefore

$$\Delta_{\rho,\varepsilon}(T)\mathbf{1}\{T \le 2\varepsilon\} \le C_K(1+\varepsilon^{-1})\mathbf{1}\{T \le 2\varepsilon\} + H_K(T)\mathbf{1}\{T \le 2\varepsilon\}.$$

Taking expectations,

$$\mathbb{E}[\Delta_{\rho,\varepsilon}(T)\mathbf{1}\{T \le 2\varepsilon\}] \le C_K(1+\varepsilon^{-1})\mathbb{P}_b(T \le 2\varepsilon) + \mathbb{E}[H_K(T)\mathbf{1}\{T \le 2\varepsilon\}].$$

The second term converges to zero by dominated convergence.

For the first term, let g_b be the marginal density of T under boundary b. The proof of Lemma 2 gives, for small t,

$$g_b(t) \le Ct^{-3/2} \exp\left(-\frac{b^2}{2t}\right).$$

Hence

$$\mathbb{P}_b(T \le 2\varepsilon) \le C \int_0^{2\varepsilon} t^{-3/2} \exp\left(-\frac{b^2}{2t}\right) dt.$$

Using the change of variables

$$x = \frac{b^2}{2t}, \qquad t = \frac{b^2}{2x}, \qquad dt = -\frac{b^2}{2x^2}\, dx,$$

we get

$$\int_0^{2\varepsilon} t^{-3/2} \exp\left(-\frac{b^2}{2t}\right) dt = C_b \int_{b^2/(4\varepsilon)}^{\infty} x^{-1/2} e^{-x}\, dx.$$

For large a,

$$\int_a^\infty x^{-1/2} e^{-x}\, dx = O(a^{-1/2} e^{-a}).$$

Therefore

$$\mathbb{P}_b(T \le 2\varepsilon) = O\left(\varepsilon^{1/2} \exp\left(-\frac{b^2}{4\varepsilon}\right)\right).$$

Thus

$$(1+\varepsilon^{-1})\mathbb{P}_b(T \le 2\varepsilon) = O\left(\varepsilon^{-1/2} \exp\left(-\frac{b^2}{4\varepsilon}\right)\right) \to 0.$$

Hence

$$\lim_{\varepsilon \downarrow 0} \mathbb{E}[\Delta_{\rho,\varepsilon}(T)\mathbf{1}\{T \le 2\varepsilon\}] = 0$$

uniformly over $\rho \in (0, b/2]$.

Combining the two regions gives

$$\limsup_{\varepsilon \downarrow 0} \mathbb{E}[\Delta_{\rho,\varepsilon}(T)] \le \mathbb{E}[r_\rho(T)].$$

Letting $\rho \downarrow 0$ yields

$$\lim_{\rho \downarrow 0} \limsup_{\varepsilon \downarrow 0} \mathbb{E}[\Delta_{\rho,\varepsilon}(T)] = 0.$$

This completes the proof of Lemma 5 and hence the proof of Theorem 6. □

Remark 5 (Uniform stochastic non-decision time)**.** *Consider $S = T + A$ with $A \sim \mathrm{Uniform}[a, a + \Delta]$, where $a \geq 0$, $\Delta > 0$, and A is independent of (V, Z, T). Uniform non-decision-time variability is a standard extension of the DDM [56]. If $|V| \leq M$ almost surely, then*

$$L_S(\lambda) = e^{-a\lambda} \frac{1 - e^{-\Delta\lambda}}{\Delta\lambda} L_T(\lambda),$$
$$\log L_S(\lambda) = -a\lambda - b\sqrt{2\lambda} - \log\lambda + \log\left(\frac{2C_0}{\Delta}\right) + O(\lambda^{-1/2}),$$

where $C_0 := \mathbb{E}[\cosh(bV)]$. The three-scale formulas in Proposition 1 consequently satisfy $\widehat{b}_n^{\mathrm{NDT}} \xrightarrow{P} b$ and $\widehat{a}_n \xrightarrow{P} a$ under (4)*. The additional $-\log\lambda$ term contributes vanishing terms to both estimates, and the relative empirical-Laplace errors still converge to zero. Here a is the lower endpoint of the non-decision-time distribution.*

The same three scales also yield

$$\widehat{C}_n^{\mathrm{NDT}} := \frac{9\lambda_n}{64} \frac{\{\widehat{L}_n^S(\lambda_n)\}^3 \widehat{L}_n^S(9\lambda_n)}{\{\widehat{L}_n^S(4\lambda_n)\}^3} \xrightarrow{P} \frac{2C_0}{\Delta}. \tag{74}$$

Indeed, the coefficients $(3, -3, 1)$ cancel both the terms proportional to λ and those proportional to $\sqrt{\lambda}$ in the logarithm of this ratio; the factor $9\lambda/64$ cancels the remaining logarithmic term. A calibration sample with the same (b, a, Δ) and a known fixed drift v_0 has $C_0 = \cosh(bv_0)$, giving the consistent estimator $\widehat{\Delta}_n = 2\cosh(\widehat{b}_n^{\mathrm{NDT}} v_0)/\widehat{C}_n^{\mathrm{NDT}}$. At zero drift this simplifies to $2/\widehat{C}_n^{\mathrm{NDT}}$.

Drift estimation with stochastic non-decision time uses the corresponding convolution law. Subtracting a leaves a uniform residual delay, while the latent response-time transform is recovered through

$$L_T(\lambda) = e^{a\lambda} \frac{\Delta\lambda}{1 - e^{-\Delta\lambda}} L_S(\lambda).$$

Thus the calibration identifies the latency parameters entering this response-time model.

B.2 Robustness to boundary heterogeneity

While we allow for heterogeneous drifts, the common-boundary assumption allows us to learn a single boundary from the aggregated data and then use it to estimate the average drift. However, if we allow for both heterogeneous drifts and arbitrary heterogeneous boundaries, learning the boundary from one observation per labeler is not possible. That said, one may still ask what our estimator converges to when the common-boundary assumption fails. This is important because it can be viewed as a robustness question. The next result shows that, in this case, the estimator actually estimates the lower endpoint of the boundary distribution. Throughout this section, we do not consider the non-decision response time.

Proposition 2. *Let (B, V) be an arbitrary random pair of boundary and drift, satisfying*

$$0 < b_- \leq B \leq b_+ < \infty, \qquad |V| \leq M < \infty \qquad \textit{almost surely},$$

where $b_- := \operatorname{ess\,inf} B$. Let T be the DDM response time conditionally on (B, V). Define the population Laplace transform $L(\lambda) := \mathbb{E}\big[e^{-\lambda T}\big]$. Then

$$-\frac{\log L(\lambda)}{\sqrt{2\lambda}} \longrightarrow b_- \qquad \textit{as } \lambda \to \infty. \tag{75}$$

Proof. Recall that

$$L(\lambda) = \mathbb{E}\left[\frac{\cosh(BV)}{\cosh\big(B\sqrt{V^2 + 2\lambda}\big)}\right].$$

We first prove the lower bound on the logarithmic rate. Similar to the derivation of (22), we have

$$L(\lambda) \leq 2e^{b_+ M} \exp\Big(-b_- \sqrt{2\lambda}\Big).$$

Therefore

$$-\log L(\lambda) \geq b_- \sqrt{2\lambda} - \log\bigl(2e^{b_+ M}\bigr)\,.$$

Dividing by $\sqrt{2\lambda}$ and taking the limit inferior yields

$$\liminf_{\lambda\to\infty} \left(-\frac{\log L(\lambda)}{\sqrt{2\lambda}}\right) \geq b_-. \tag{76}$$

We next prove the reverse inequality. Fix $\varepsilon > 0$. By the definition of the essential infimum,

$$p_\varepsilon := \mathbb{P}(B \leq b_- + \varepsilon) > 0.$$

Now, using this, and similar to the derivation of (21), we have Hence

$$L(\lambda) \geq p_\varepsilon \exp\Bigl(-(b_- + \varepsilon)\sqrt{M^2 + 2\lambda}\Bigr)\,.$$

Therefore

$$-\log L(\lambda) \leq (b_- + \varepsilon)\sqrt{M^2 + 2\lambda} - \log p_\varepsilon.$$

Dividing by $\sqrt{2\lambda}$ and taking the limit superior gives

$$\limsup_{\lambda\to\infty} \left(-\frac{\log L(\lambda)}{\sqrt{2\lambda}}\right) \leq b_- + \varepsilon. \tag{77}$$

Since $\varepsilon > 0$ was arbitrary,

$$\limsup_{\lambda\to\infty} \left(-\frac{\log L(\lambda)}{\sqrt{2\lambda}}\right) \leq b_-. \tag{78}$$

Combining (76) and (78) proves (75). □

Proposition 2 implies that arbitrary unobserved boundary heterogeneity is not harmless. That said, this also means that when the boundary is not homogeneous but still lies in a small interval, our estimator returns an approximately accurate estimate of it. One can derive bounds on the final estimation error as a function of $b_+ - b_-$.

One final remark is that when boundary heterogeneity is associated with observed strata, such as task type, annotation interface, demographic pool, or any other observed grouping variable, we can group users accordingly, learn one boundary for each group, and again apply the results derived under the homogeneous-boundary case.

B.3 Relaxing the bounded drifts assumption

In this section, we relax the bounded-drift assumption to require only bounded exponential moments. This relaxation ensures our estimators provably accommodate Gaussian drift priors. Our contributions in this section are summarized as follows:

- In Theorem 2 and Theorem 3, we relax the almost everywhere (a.e.) boundedness assumption, $|V| \leq M$ a.e., to the moment condition $\mathbb{E}\bigl[(1 + V^4)\cosh(bV)\bigr] < \infty$.
- In Theorem 5, we similarly relax the assumption $\|\psi_i\| \leq L$ and $\|\theta_i\| \leq R$ a.e., to $\mathbb{E}\bigl[\bigl(1 + \|\psi_i\|^2 + \|\psi_i\|^4\|\theta_i\|^4\bigr)\cosh\bigl(b\psi_i^\top\theta_i\bigr)\bigr] < \infty$.

Our first result recovers consistent boundary estimate under the relaxed assumption.

Proposition 3 (Empirical boundary consistency under exponential moments)**.** *Let $(V_i, T_i)_{i=1}^n$ be i.i.d. as follows: $V_i \sim F^*$, and conditional on $V_i = v$, T_i is the first-passage time of a DDM with common boundary $b > 0$ and drift v. Conditional on $(V_i)_{i=1}^n$, the DDM paths are independent. Assume*

$$C_0 = \mathbb{E}[\cosh(bV)] < \infty. \tag{79}$$

Then,

$$-\frac{\log L_b(\lambda)}{\sqrt{2\lambda}} \to b. \tag{80}$$

Furthermore, suppose $\{\lambda_n\}_n$ *satisfies* (4)*. Then,*

$$\frac{\widehat{L}_n(\lambda_n)}{L_b(\lambda_n)} \xrightarrow{P} 1, \quad -\frac{\log \widehat{L}_n(\lambda_n)}{\sqrt{2\lambda_n}} \xrightarrow{P} b. \tag{81}$$

The same conclusion holds for the Richardson estimator $\widehat{B}_n$ *in* (5)*.*

Proof. We first show

$$e^{b\sqrt{2\lambda}} L_b(\lambda) \to 2C_0 \qquad \text{as } \lambda \to \infty. \tag{82}$$

For each fixed $v \in \mathbb{R}$,

$$\sqrt{v^2 + 2\lambda} - \sqrt{2\lambda} \to 0 \qquad \text{as } \lambda \to \infty.$$

Using $\cosh(x) = (e^x + e^{-x})/2$, we obtain the pointwise limit

$$\frac{e^{b\sqrt{2\lambda}} \cosh(bv)}{\cosh\big(b\sqrt{v^2 + 2\lambda}\big)} \to 2\cosh(bv).$$

For domination, since $\cosh(y) \geq e^y/2$ for $y \geq 0$,

$$\frac{e^{b\sqrt{2\lambda}} \cosh(bv)}{\cosh\big(b\sqrt{v^2 + 2\lambda}\big)} \leq 2\cosh(bv) \exp\Big(-b\{\sqrt{v^2 + 2\lambda} - \sqrt{2\lambda}\}\Big) \leq 2\cosh(bv).$$

The dominating function $2\cosh(bV)$ is integrable by (79). Dominated convergence therefore gives (82). Taking logarithms in (82) yields

$$\log L_b(\lambda) = -b\sqrt{2\lambda} + \log(2C_0) + o(1).$$

Dividing by $\sqrt{2\lambda}$ proves (80).

Since $0 \leq e^{-\lambda T} \leq 1$,

$$\mathrm{Var}\big(e^{-\lambda T}\big) \leq \mathbb{E}\big[e^{-2\lambda T}\big] = L_b(2\lambda).$$

Therefore

$$\mathbb{E}\left[\left(\frac{\widehat{L}_n(\lambda_n)}{L_b(\lambda_n)} - 1\right)^2\right] = \frac{\mathrm{Var}(e^{-\lambda_n T})}{nL_b(\lambda_n)^2} \leq \frac{L_b(2\lambda_n)}{nL_b(\lambda_n)^2}.$$

By (82),

$$L_b(2\lambda) = O\Big(e^{-b\sqrt{4\lambda}}\Big), \qquad L_b(\lambda)^2 = \Theta\Big(e^{-2b\sqrt{2\lambda}}\Big).$$

Hence

$$\frac{L_b(2\lambda_n)}{nL_b(\lambda_n)^2} = O\left(\frac{\exp\big(2b(\sqrt{2}-1)\sqrt{\lambda_n}\big)}{n}\right). \tag{83}$$

Condition (4) implies that the right-hand side of (83) converges to zero. This proves (81). The Richardson estimator uses the same argument at scales λ_n and $4\lambda_n$, and $2\sqrt{\lambda_n} = o(\log n)$ still holds. Thus $\widehat{B}_n$ also converges in probability to b. □

It is worth noting that the scalar exponential-moment condition includes Gaussian drifts. Indeed, if $V \sim N(m, \sigma^2)$, then

$$\mathbb{E}[\cosh(bV)] = \frac{1}{2}\left(e^{bm + b^2\sigma^2/2} + e^{-bm + b^2\sigma^2/2}\right) < \infty.$$

This result implies that the estimator $\widehat{B}_n$, which we developed earlier, is consistent under the lighter moment-type assumption (79), rather than under the boundedness assumption. That said, to establish the second part of Theorem 3, which shows that this estimator enjoys a better rate than the one-scale estimator, we need a slightly stronger assumption, which still holds for Gaussian drifts. The next proposition formalizes this.

Proposition 4. *Under the premise of Proposition 3, assume*

$$\mathbb{E}\big[(1+V^4)\cosh(bV)\big] < \infty. \tag{84}$$

Then, (6) *holds, meaning that the population version of $\widehat{B}_n$ enjoys a faster convergence rate than the one-scale estimator.*

Proof. Recall from the proof of Theorem 3 that

$$C_0 := \mathbb{E}[\cosh(bV)], \qquad C_2 := \mathbb{E}[V^2\cosh(bV)].$$

Let

$$s := \sqrt{2\lambda}, \qquad \delta_s(v) := \sqrt{s^2+v^2} - s.$$

Then

$$\delta_s(v) = \frac{v^2}{\sqrt{s^2+v^2}+s}, \qquad 0 \le \delta_s(v) \le \frac{v^2}{2s}.$$

Also,

$$\frac{1}{\cosh(b\sqrt{s^2+v^2})} = 2e^{-bs}e^{-b\delta_s(v)}\left(1+\rho_s(v)\right),$$

where

$$|\rho_s(v)| \le e^{-2b\sqrt{s^2+v^2}} \le e^{-2bs}.$$

Therefore

$$L_b(\lambda) = 2e^{-bs}\mathbb{E}\Big[\cosh(bV)e^{-b\delta_s(V)}\Big] + O(e^{-3bs}C_0). \tag{85}$$

We now expand the remaining expectation. Since $0 \le \delta_s(V) \le V^2/(2s)$ and

$$\left|e^{-b\delta_s(V)} - 1 + b\delta_s(V)\right| \le \frac{b^2\delta_s(V)^2}{2} \le \frac{b^2V^4}{8s^2},$$

we have

$$\mathbb{E}\Big[\cosh(bV)e^{-b\delta_s(V)}\Big] = C_0 - b\mathbb{E}[\cosh(bV)\delta_s(V)] + O(s^{-2}). \tag{86}$$

Next,

$$0 \le \frac{V^2}{2s} - \delta_s(V) = \frac{V^2(\sqrt{s^2+V^2}-s)}{2s(\sqrt{s^2+V^2}+s)} \le \frac{V^4}{8s^3}.$$

Thus

$$\mathbb{E}[\cosh(bV)\delta_s(V)] = \frac{C_2}{2s} + O(s^{-3}). \tag{87}$$

Combining (85), (86), and (87) gives

$$L_b(\lambda) = 2e^{-bs}\left(C_0 - \frac{bC_2}{2s} + O(s^{-2})\right).$$

Taking logarithms,

$$\begin{aligned}\log L_b(\lambda) &= -bs + \log(2C_0) + \log\left(1 - \frac{bC_2}{2C_0 s} + O(s^{-2})\right)\\ &= -bs + \log(2C_0) - \frac{bC_2}{2C_0 s} + O(s^{-2}).\end{aligned}$$

Since $s = \sqrt{2\lambda}$, implies

$$\log L_b(\lambda) = -b\sqrt{2\lambda} + \log(2C_0) - \frac{bC_2}{2C_0\sqrt{2\lambda}} + O(\lambda^{-1}). \tag{88}$$

Let

$$c_0 := \log(2C_0), \qquad c_1 := -\frac{bC_2}{2C_0}.$$

Then (88) can be written as

$$\log L_b(\lambda) = -bs + c_0 + \frac{c_1}{s} + O(s^{-2}). \tag{89}$$

At scale 4λ, the corresponding value of s is $2s$, and hence

$$\log L_b(4\lambda) = -2bs + c_0 + \frac{c_1}{2s} + O(s^{-2}). \tag{90}$$

Subtracting (90) from (89) gives

$$\log L_b(\lambda) - \log L_b(4\lambda) = bs + \frac{c_1}{2s} + O(s^{-2}).$$

Dividing by $s = \sqrt{2\lambda}$ proves the rate for the Richardson estimator. Finally,

$$-\frac{\log L_b(\lambda)}{s} - b = -\frac{c_0}{s} - \frac{c_1}{s^2} + O(s^{-3}). \tag{91}$$

Because $C_0 = \mathbb{E}[\cosh(bV)] \geq 1$, we have $c_0 = \log(2C_0) > 0$. Hence the leading term in (91) is nonzero, and the other rate follows. □

Finally, we can similarly relax the boundedness assumption in the linear model.

Proposition 5 (Linear preference consistency under integrability conditions)**.** *Consider the setting of linear preferences in Section 4. Assume*

$$\mathbb{E}\big[\big(1 + \|\psi_i\|^2 + \|\psi_i\|^4\|\theta_i\|^4\big)\cosh\big(b\psi_i^\top\theta_i\big)\big] < \infty. \tag{92}$$

Then $\widehat{\theta}_n \xrightarrow{P} \theta^\star$.

Proof. The proof generalizes the proof of Theorem 5, provided in Appendix appendix A.5. By the weak law of large numbers and $\mathbb{E}\|\psi_i\|^2 < \infty$,

$$\widehat{Q}_n \xrightarrow{P} Q. \tag{93}$$

Therefore, we have $\widehat{Q}_n^{-1} \xrightarrow{P} Q^{-1}$.

Next, similarly, we consider the decomposition $\widehat{m}_n = S_n + R_n$, where, recall that,

$$S_n = \frac{1}{n}\sum_{i=1}^n \psi_i Z_i w_b(T_i), \qquad R_n = \frac{1}{n}\sum_{i=1}^n \psi_i Z_i \left\{w_{\widehat{b}_n}(T_i) - w_b(T_i)\right\}.$$

For the main term, condition on (ψ_i, θ_i). The scalar unbiasedness identity gives

$$\mathbb{E}[Z_i w_b(T_i) \mid \psi_i, \theta_i] = \psi_i^\top\theta_i.$$

Therefore

$$\mathbb{E}[\psi_i Z_i w_b(T_i)] = \mathbb{E}[\psi_i\psi_i^\top\theta_i] = Q\theta^\star.$$

The moment assumption implies $\mathbb{E}[\|\psi_i\|^2\|\theta_i\|] < \infty$, which ensures integrability of $\psi_i\psi_i^\top\theta_i$. Furthermore, $\phi_i Z_i w_b(T_i)$ is absolute integrable due to

$$\mathbb{E}\left[\|\psi_i\|\mathbb{E}[|w_b(T_i)| \mid \psi_i, \theta_i]\right] \lesssim \mathbb{E}\left[\|\psi_i\|\cosh(b\psi_i^\top\theta_i)\right] < \infty.$$

So the weak law of large numbers implies

$$S_n \xrightarrow{P} Q\theta^\star. \tag{94}$$

It remains to show $R_n \xrightarrow{P} 0$. Fix $\eta > 0$ and let

$$E_{n,\eta} := \{|\widehat{b}_n - b| \le \eta\}.$$

By Proposition 4, $\widehat{b}_n \xrightarrow{P} b$, so $\mathbb{P}(E_{n,\eta}) \to 1$. On $E_{n,\eta}$,

$$\|R_n\| \le \frac{1}{n}\sum_{i=1}^{n} \|\psi_i\| \sup_{|\alpha-b|\le\eta} |w_\alpha(T_i) - w_b(T_i)|.$$

The expectation of the right-hand side converges to zero as $\eta \downarrow 0$ due to

$$\mathbb{E}\left[\|\psi_i\| \sup_{|\alpha-b|\le\eta} |w_\alpha(T_i) - w_b(T_i)|\right] \le \left(\mathbb{E}\big[\|\psi_i\|^2\big]\,\mathbb{E}\left[\sup_{|\alpha-b|\le\eta} |w_\alpha(T_i) - w_b(T_i)|^2\right]\right)^{-1/2} \downarrow 0$$

by Lemma 2. Therefore

$$R_n \xrightarrow{P} 0. \tag{95}$$

Combining (93), (94), and (95),

$$\widehat{\theta}_n = \widehat{Q}_n^{-1}\widehat{m}_n \xrightarrow{P} Q^{-1}Q\theta^\star = \theta^\star,$$

which completes the proof. □

If ψ_i is bounded and θ_i is Gaussian, then (92) holds automatically. Hence Gaussian latent preferences do not require truncation for the boundary-identification argument.

B.4 A finite-dimensional realization of the distortion construction

We give an explicit linear-utility realization of the construction in Gölz et al. [28, Theorem 3], with the dimension and alternative set fixed before taking the sample-size limit.

Proposition 6 (Choice-only distortion with finitely many alternatives)**.** *Fix $\beta > 0$ and $\varepsilon > 0$. There exist a finite integer $m \ge 2$ and an alternative set $\mathcal{X} = \{e_1, \dots, e_m\} \subset \mathbb{R}^m$, where e_j is the j-th standard basis vector, with the following property. For every possibly randomized algorithm $\mathcal{A}$ using only one binary comparison from each of n independent users, there is a population distribution G^* supported on two vectors in $[0,1]^m$ such that the linear utilities $u_i(x) = x^\top\theta_i$, $\theta_i \sim G^*$, satisfy*

$$\limsup_{n\to\infty} \frac{\max_{x\in\mathcal{X}} \mathrm{AvgUtil}(x)}{\mathrm{AvgUtil}_n(\mathcal{A})} \ge \frac{\beta}{2}\frac{1+e^{-\beta}}{1-e^{-\beta}} - \varepsilon. \tag{96}$$

Comparisons follow the DDM with common boundary $b = \beta/2$, and comparison pairs are drawn independently of user parameters from a fixed distribution on distinct alternatives. Both m and G^* remain fixed as $n \to \infty$.

Proof. Let $m \ge 2$ and $0 < c < 1$, to be chosen below, and define

$$p_c := \frac{\tanh(\beta c/2)}{\tanh(\beta/2) + \tanh(\beta c/2)}.$$

For each possible distinguished alternative $a = e_j$, let G_a assign probability p_c to $\theta^{(1)} = e_j$ and probability $1 - p_c$ to $\theta^{(2)} = c\sum_{k\ne j} e_k$. The corresponding utility vectors have value 1 at a and 0 elsewhere in the first case, and value 0 at a and c elsewhere in the second. For any other alternative y, the pooled probability of choosing a over y is

$$\frac{p_c}{1+e^{-\beta}} + \frac{1-p_c}{1+e^{\beta c}} = \frac{1}{2}.$$

Every comparison between two alternatives other than a also has choice probability $1/2$. Since each independent user contributes only one comparison, the entire choice-data distribution is identical for all m populations G_a.

The average utilities under G_a are $\mathrm{AvgUtil}(a) = p_c$ and $\mathrm{AvgUtil}(y) = (1 - p_c)c$ for $y \neq a$. Their ratio is

$$r_c := \frac{p_c}{(1-p_c)c} = \frac{\tanh(\beta c/2)}{c\tanh(\beta/2)} > 1,$$

where the strict inequality follows from strict concavity of $\tanh$ on $(0, \infty)$. Thus a is the unique average-utility maximizer.

Let $q_{n,j}$ denote the probability that $\mathcal{A}$ selects e_j, averaging over the observed choices and the algorithm's randomization. These probabilities are the same under every G_a and sum to one. For each n, some j satisfies $q_{n,j} \leq 1/m$; since there are finitely many alternatives, a fixed j satisfies this inequality along an infinite subsequence. Choose $G^* = G_{e_j}$. Along that subsequence,

$$\begin{aligned}\mathrm{AvgUtil}_n(\mathcal{A}) &= p_c q_{n,j} + (1-p_c)c(1-q_{n,j}) \\ &\leq \frac{p_c}{m} + \frac{m-1}{m}(1-p_c)c.\end{aligned}$$

Consequently,

$$\limsup_{n\to\infty} \frac{\max_{x\in\mathcal{X}} \mathrm{AvgUtil}(x)}{\mathrm{AvgUtil}_n(\mathcal{A})} \geq \frac{mr_c}{r_c + m - 1}.$$

Finally,

$$\lim_{c\downarrow 0} r_c = \frac{\beta}{2\tanh(\beta/2)} = \frac{\beta}{2}\frac{1+e^{-\beta}}{1-e^{-\beta}}, \qquad \lim_{m\to\infty} \frac{mr_c}{r_c+m-1} = r_c.$$

Choose $c > 0$ sufficiently small and then a finite m sufficiently large that the last lower bound is within ε of the claimed value. These choices precede the limit in n, establishing (96). □

For each fixed instance, the feature and parameter vectors are bounded. The comparison vectors span $\mathbf{1}^\perp$, and a design assigning positive probability to every distinct pair has positive-definite covariance on this subspace. Utility differences can therefore be represented in a full-rank $(m-1)$-dimensional basis. The common additive utility coordinate does not affect the maximizing alternative; the distortion ratios above use the original nonnegative utilities.

C Detailed Experimental Settings

This appendix provides the implementation details, hyperparameter settings, and compute cost that were deferred from Section 5.

C.1 Common setup

Boundary estimator. Throughout, we use the bias-corrected boundary estimator $\widetilde{B}_n$ of Eq. (5) with $\lambda_n = (\log n)^{3/2}$.

Weight function evaluation. The ratio-of-series weight $w_\beta(t)$ of Eq. (2) is evaluated by truncating both series at $K = 100$ terms. The truncation error decays super-exponentially in K: for $t \leq T$ and $\beta \geq b_-$, the k-th term is bounded by $(2k+1)e^{-(2k+1)^2 b_-^2/(2T)}$, which for the parameter ranges used in our experiments gives machine-precision error well before $K = 100$.

First-passage-time sampling. Synthetic response times are drawn from the conditional density $f(t \mid v, b) = f_0(t; b)\exp(-v^2 t/2)\cosh(bv)$ by inverse-CDF sampling on a precomputed cache; the real-data experiment uses the recorded τ_i directly. We tabulate f on the product of an N_v-point uniform drift grid and a 998-point time grid concentrated near the mode of f_0 ($\Delta t \approx 5 \cdot 10^{-5}$ on $[10^{-4}, 10^{-2}]$, coarsening to $\Delta t \approx 6.5 \cdot 10^{-2}$ in the tail) and truncated at $T_{\max} = 20$; the leading large-t asymptote $f_0(t; b) \sim (\pi/2b^2)\exp(-\pi^2 t/(8b^2))$ together with the exponential tilt $\exp(-v^2 t/2)$ bounds the truncated tail mass by 10^{-7} for all $|v| \leq V_{\max}$ and $b \geq 1.25$ used here. The CDF is built row-wise by the trapezoidal rule and renormalized so that $F(T_{\max} \mid v, b) = 1$ exactly; sampling snaps each drift to its nearest v-grid point and linearly interpolates in t, with bias $O(\Delta v)$ and $O((\Delta t)^2)$

respectively. We use $N_v = 500$ over the prior support for the basic experiments and $N_v = 801$ over $[-V_{\max}, V_{\max}]$ for the contextual experiments, where $V_{\max} = \sum_{k=1}^{d} \max\big(|\theta_k^\star - 6\sigma_\theta|, |\theta_k^\star + 6\sigma_\theta|\big) = 12.80$ matches the $\pm 6\sigma_\theta$ truncation of the priors. The cache is built once per scenario and shared across all repetitions; in the contextual setting, if a realized drift ever exceeds $V_{\max}$ during a rep-batch the cache is rebuilt with $V_{\max} \leftarrow \max(1.5\, V_{\max},\, 1.1 \max_i |V_i|)$, an event that is a deterministic function of the seeded (ψ_i, θ_i) draws, so the rebuild sequence is identical across re-runs at fixed seed, dtype, and device.

Replications and randomness. Every MSE and cosine-similarity curve is the Monte-Carlo average over $R = 50$ independent replications with independent random seeds. For the synthetic experiments, confidence regions are generated pointwise with half-width $1.96 \cdot \mathrm{std}/\sqrt{R}$; the linear plotting script uses the same formula with the squared-error array. The shaded bands in Figure 3 are 95% bootstrap confidence bands over the R replications.

Compute. All experiments were run on a single NVIDIA A100-80GB GPU within 10 minutes.

C.2 Synthetic tabular setting

We fix $\mu^\star = \mathbb{E}_{F^*}[V] = 0.25$ and $b^\star = 1.25$.

Prior families.

- *Uniform:* $V \sim \mathrm{Uniform}[-0.25, 0.75]$, which has mean $\mu^\star$ and support width 1.
- *Beta:* $V = \mu^\star - \frac{2}{7} + \mathrm{Beta}(2, 5)$, a centered Beta rescaled to unit width, yielding a unimodal skewed density.

Metric. The Monte-Carlo MSE $\mathbb{E}[(\widehat{\mu}_n - \mu^\star)^2]$, averaged over $R = 50$ replications.

Sample-size grid. $n \in \{10^3, 10^4, 10^5, 10^6\}$.

C.3 Synthetic linear contextual setting

We fix $d = 4$, $\theta^\star = (0.25, -0.15, 0.10, -0.30)$, $\sigma_\theta^2 = 0.25$, and $b^\star = 1.25$. Contexts are $\psi_i \sim \mathrm{Uniform}[-1, 1]^d$ i.i.d., and each coordinate of θ_i is drawn i.i.d. from one of the following priors centered at the corresponding coordinate of $\theta^\star$ with variance σ_θ^2:

- *Gaussian:* $\theta_i^{(k)} \sim \mathcal{N}(\theta_k^\star, \sigma_\theta^2)$, truncated to $\pm 6\sigma_\theta$.
- *Uniform:* $\theta_i^{(k)} \sim \mathrm{Uniform}[\theta_k^\star - \sigma_\theta\sqrt{3},\, \theta_k^\star + \sigma_\theta\sqrt{3}]$.
- *Beta:* $\theta_i^{(k)} = \theta_k^\star + \sigma_\theta\, \tilde{b}_k$ for a mean-centered $\mathrm{Beta}(2, 5)$ variable rescaled to unit variance.
- *Laplace:* $\theta_i^{(k)} \sim \mathrm{Laplace}(\theta_k^\star, \sigma_\theta/\sqrt{2})$, truncated to $\pm 6\sigma_\theta$.

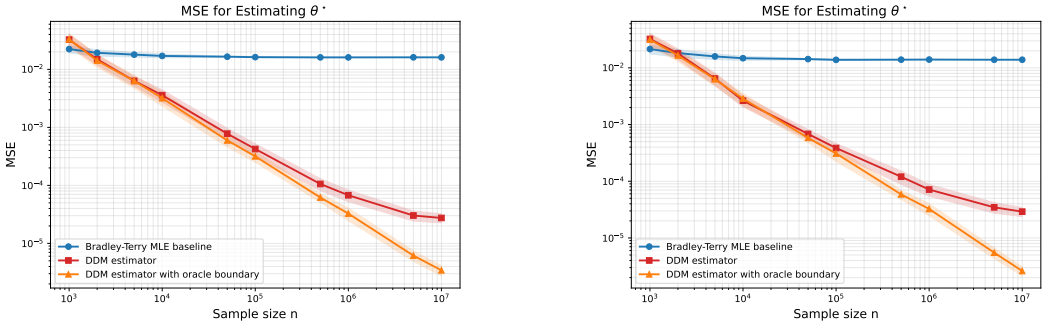
Metric. The Monte-Carlo MSE $\mathbb{E}[\|\widehat{\theta}_n - \theta^\star\|_2^2]$, averaged over $R = 50$ replications.

Sample-size grid. $n \in \{10^3, 10^4, 10^5, 10^6, 10^7\}$.

C.4 Real-data intertemporal choice

Source. We use the publicly available dataset of Amasino et al. [3] as reanalyzed by Echenique et al. [19]. Each trial presents a binary choice between a smaller-sooner reward $s_r \in \{0.5, \ldots, 9.5\}$ dollars paid today ($s_d = 0$) and a larger-later reward $\ell_r = 10$ dollars paid after delay $\ell_d \in \{1, 7, 15, 30, 90, 180, 365\}$ days. The dataset records the binary choice $Z_i \in \{+1, -1\}$ (encoded $+1$ for larger-later) and the raw reaction time $\tau_i > 0$.

Preprocessing. We drop trials with missing choices or response times, and we drop two participants whose retained trials contain only one choice class (so that the subject-specific Bradley–Terry MLE is degenerate). The resulting replication sample pools $n_{\max} = 13{,}793$ valid trials from $S = 98$ participants (out of 100 enrolled).



(a) Beta. (b) Laplace.

Figure 4: **Additional results of the linear contextual setting.** Monte-Carlo MSE of $\widehat{\theta}_n$ versus sample size n for two other priors on $\theta_i \in \mathbb{R}^4$. The DDM estimator again tracks the oracle and uniformly dominates the baseline. Shaded bands denote pointwise 95% confidence intervals over $R = 50$ replications.

Feature encoding. Following Echenique et al. [19], each trial is encoded by the normalized comparison-feature vector $\psi_i = \big((\ell_r - s_r)/9.5,\ -(\ell_d - s_d)/365\big) \in \mathbb{R}^2$. The first coordinate is the scaled monetary gap and the second is the negated scaled delay gap.

Regularization. In 2-D feature space, it is possible that some participants' choices are perfectly linearly separable, which results in divergence of unregularized Bradley–Terry MLE. Therefore, we add a ℓ_2-regularization with penalty parameter $\lambda = 0.1$ to all Bradley–Terry MLE losses in the real data setting.

Subsampling. For each n in the grid $\{10^2, 250, 500, 10^3, 2000, 5000\}$, we draw $R = 50$ random subsamples of size n with replacement from the pooled trials. Each estimator is fit on each subsample.

Metric. Because the common boundary b is unidentifiable from choices alone, only the direction of $\theta^\star$ is comparable across estimators. We therefore report

$$\mathrm{CosSim}(\widehat{\theta}, \theta^\star) \;=\; \frac{\widehat{\theta}^\top \theta^\star}{\|\widehat{\theta}\|_2 \|\theta^\star\|_2} \in [-1, 1],$$

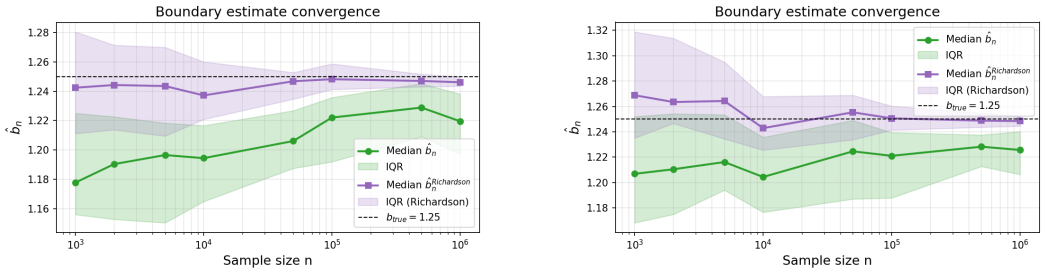
averaged over the $R = 50$ subsamples of size n, with 95% bootstrap bands shown as the shaded region in Figure 3.

On the DDM assumption. The real-data experiment is not an identification claim: human decision-making in intertemporal choice is only approximately modeled by the DDM. The fact that the DDM estimator nonetheless aligns more closely with the subject-level target direction than the pooled Bradley–Terry baseline, and continues to improve where the baseline plateaus, is evidence that response times carry information about the aggregate preference direction even when the generative model is potentially misspecified.

C.5 Ablation: Richardson extrapolation for boundary estimation

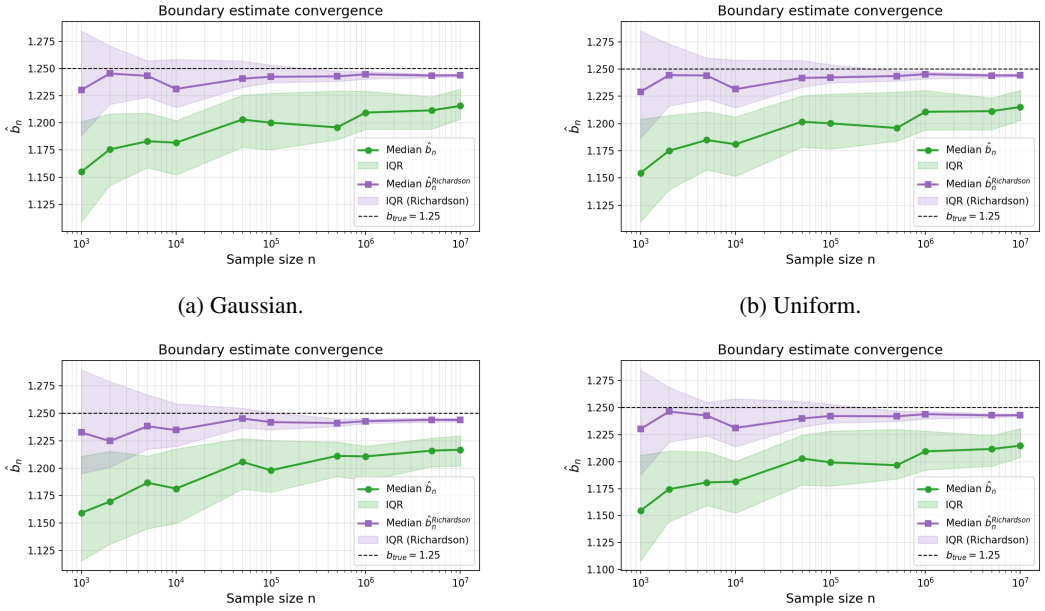
We directly compare the two-scale Richardson-extrapolated estimator $\widetilde{B}_n$ of Eq. (5) against the uncorrected one-scale estimator $\widehat{b}_n := -\frac{\log \widehat{L}_n(\lambda_n)}{\sqrt{2\lambda_n}}$, with both estimators using the same Laplace-tail rate $\lambda_n = (\log n)^{3/2}$ and the same response-time samples.

Setup. We reuse the synthetic data-generating processes of Sections 5.1 and Appendix C: tabular data with $b^\star = 1.25$ and the prior on V varied across $\{\text{Uniform}, \text{Beta}, \text{Gaussian}, \text{Laplace}\}$, and contextual data with $b^\star = 1.25$ and the per-coordinate prior on $\theta_i \in \mathbb{R}^4$ varied across the same four families. For each n we report the median and interquartile range (IQR) over $R = 10$ replications.



(a) Uniform. (b) Beta.

Figure 5: **Tabular setting.** Median and IQR (over $R = 10$ replications) of the one-scale estimator $\widehat{b}_n$ (green) and the Richardson-extrapolated estimator $\widetilde{B}_n$ (purple) as a function of sample size n, for each of the two priors on the latent drift V. The dashed line marks $b^\star = 1.25$. The Richardson correction removes the $O(\lambda_n^{-1/2})$ bias of $\widehat{b}_n$ across all priors.



(a) Gaussian. (b) Uniform.

(c) Beta. (d) Laplace.

Figure 6: **Linear contextual setting.** Median and IQR (over $R = 10$ replications) of $\widehat{b}_n$ (green) and $\widetilde{B}_n$ (purple) versus sample size n, for each of the four per-coordinate priors on $\theta_i \in \mathbb{R}^4$. The dashed line marks $b^\star = 1.25$. The Richardson correction is also effective in the contextual setting, where the latent drifts $v_i = \psi_i^\top \theta_i$ are heterogeneous through both ψ_i and θ_i.

Results. Across all priors and both settings (Figures 5 and 6), the Richardson-extrapolated estimator $\widetilde{B}_n$ achieves smaller error and faster convergence to $b^\star = 1.25$ than the uncorrected one-scale estimator $\widehat{b}_n$. This directly substantiates the efficiency-gain of Theorem 3.

Prior	Estimator	$n = 10^3$	$n = 5 \cdot 10^3$	$n = 10^5$	$n = 10^7$
Gaussian	Bradley–Terry MLE	0.0229	0.0181	0.0161	0.0162
	Neyman-orthogonal RT loss	**0.0177**	0.0109	0.0081	0.0080
	DDM	0.0311	**0.0079**	**0.00042**	**0.000027**
Uniform	Bradley–Terry MLE	0.0237	0.0193	0.0174	0.0174
	Neyman-orthogonal RT loss	**0.0178**	0.0118	0.0089	0.0087
	DDM	0.0300	**0.0081**	**0.00040**	**0.000025**
Beta	Bradley–Terry MLE	0.0255	0.0176	0.0161	0.0161
	Neyman-orthogonal RT loss	**0.0206**	0.0102	0.0082	0.0080
	DDM	0.0270	**0.0065**	**0.00045**	**0.000028**
Laplace	Bradley–Terry MLE	0.0199	0.0158	0.0139	0.0139
	Neyman-orthogonal RT loss	**0.0157**	0.0094	0.0068	0.0067
	DDM	0.0301	**0.0072**	**0.00042**	**0.000032**
Bimodal	Bradley–Terry MLE	0.0253	0.0207	0.0183	0.0183
	Neyman-orthogonal RT loss	**0.0196**	0.0124	0.0094	0.0093
	DDM	0.0353	**0.0074**	**0.00041**	**0.000019**

Table 1: **Response-time baseline comparison.** Mean squared error $\mathbb{E}[\|\widehat{\theta} - \theta^\star\|_2^2]$ in the heterogeneous contextual model. The Neyman-orthogonal estimator uses the same estimated boundary as our DDM estimator. Cross-fitting and data reuse agree to within 10^{-4}. Boldface marks the lowest MSE for each prior and sample size.

C.6 Comparison with a Neyman-orthogonal response-time loss

The Neyman-orthogonal response-time loss of Sawarni et al. [60] provides a complementary approach to controlling nuisance-estimation error in a homogeneous reward model. We compare this loss with our estimator under heterogeneous preferences, using the same estimated boundary and both cross-fitting and data reuse.

Setup. We compare our estimator with the Neyman-orthogonal response-time loss of Sawarni et al. [60] on the contextual synthetic benchmark and the intertemporal-choice data of Section 5.2. We supply the Neyman-orthogonal estimator with our estimated boundary $\widetilde{B}_n$ and implement their Meta-Algorithm 1 with the EZ plug-in nuisance estimates from their Appendix A.6. We evaluate both two-fold cross-fitting and data reuse for nuisance estimation. As an implementation check, the estimator recovers the true parameter on data generated from its homogeneous working model.

Synthetic comparison. We use the contextual model of Section 5.1, with Gaussian, Uniform, Beta, Laplace, and Bimodal preference priors. For the Bimodal prior, each coordinate takes the values $\theta_k^\star - \sigma_\theta$ and $\theta_k^\star + \sigma_\theta$ with equal probability. Table 1 reports the MSE for $n \in \{10^3, 5 \cdot 10^3, 10^5, 10^7\}$. The two nuisance-estimation schemes differ by less than 10^{-4} in these experiments.

Results. The Neyman-orthogonal loss reduces the large-sample MSE of the choice-only estimator by approximately a factor of two. At $n = 10^3$, it has the smallest MSE for every prior. By $n = 5 \cdot 10^3$, the DDM estimator has lower MSE for all five priors, and its error continues to decrease over the reported sample sizes. At $n = 10^7$, its MSE is approximately 200–500 times smaller than that of the Neyman-orthogonal estimator. The plateau of the latter occurs under both nuisance-estimation schemes. Its homogeneous working model targets a pseudo-true parameter under heterogeneous preferences, whereas our pseudo-outcome construction targets $\mathbb{E}[\theta_i]$.

Intertemporal-choice comparison. Table 2 gives the cosine similarity to the subject-level average target on the dataset of Amasino et al. [3]. Both versions of the Neyman-orthogonal estimator improve on the pooled Bradley–Terry MLE, attaining a cosine similarity of 0.9942, while the DDM estimator attains 0.9983.

Estimator	Cosine similarity	Difference from DDM
Bradley–Terry MLE	0.9936	-0.0047
Neyman-orthogonal RT loss (data reuse)	0.9942	-0.0041
Neyman-orthogonal RT loss (cross-fitting)	0.9942	-0.0041
DDM	**0.9983**	—

Table 2: **Intertemporal-choice baseline comparison.** Cosine similarity to the subject-level average preference vector, using the target and feature encoding of Appendix C.4. Boldface marks the highest cosine similarity.

Study	Bradley–Terry cosine similarity	DDM cosine similarity
Lange & Eggert (2014) [37]	0.9503	**0.9979**
Jones et al. (2017) [32]	0.9984	**1.0000**
Yang et al. (2022) [73]	0.9989	0.9993
Kable et al. (2017) [33]	0.9993	0.9991
Schwenke et al. (2022) [62]	0.9986	0.9978
Calluso et al. (2019) [10]	**0.9965**	0.9929

Table 3: **Additional intertemporal-choice studies.** Cosine similarity to the participant-average preference target for the six eligible studies in the database of Pongratz and Schoemann [50]. Decision boundaries are estimated separately for each study. Boldface marks the better-performing estimator when the difference is statistically distinguishable.

C.7 Additional intertemporal-choice datasets

Data and study selection. We evaluate the estimators on the intertemporal-choice database of Pongratz and Schoemann [50], which contains 100 studies from 87 publications, $11{,}852$ participants, and $1{,}172{,}644$ trials, including the dataset of Amasino et al. [3]. Because decision boundaries are study-specific, we select 6 out of the 100 studies based on first-principled quality screening: non-adaptive elicitation, sub-second RT resolution, no short response deadline, ≥ 50 participants, $\geq 10,000$ trials, stable participant-average target, common-boundary assumption not violated by design (no adaptive stimulus selection, no multi-time-scale delays, and no systematical difference in decision caution). This screen is design-based and does not reference the target or estimator output. We estimate the boundary separately within each study and compare the pooled estimators with the participant-average target.

Results. Table 3 shows that of the 6 qualifying studies, our estimator is better on two, statistically indistinguishable on three, and worse on one. The largest gain (cosine $0.950 \to 0.998$) occurs in Lange & Eggert (2014) [37] where the baseline exhibits a clear aggregation bias floor, while the remaining five studies show negligible differences in CosSim, likely due to the high BT baseline scores indicating high homogeneity.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract and introduction accurately reflect the paper's contributions and scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of the work in Section 6.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provide full assumptions in the theorem statements and complete proofs in appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We disclose all experiment setups and include the codes in supplementary materials.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide the codes and data sources with sufficient instructions to reproduce the experiments.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We specify all test details in Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We report 0.95 confidence regions in all experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide the computer resources used to run the experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We conform with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss both positive and negative societal impacts in Section 6.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We do not release data or models that have any risk for misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We credit the creators of assets used in the paper. URL: `https://osf.io/2p3dj/overview`

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: We do not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: We do not collect new human-subject data. We use a public dataset from Amasino et al. [3]. We follow the access terms of the released dataset and do not attempt to re-identify participants.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: We do not collect new human-subject data. We use a public dataset from Amasino et al. [3]. We follow the access terms of the released dataset and do not attempt to re-identify participants.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: LLMs were mainly used for paper editing and code assistance. That said, we also used the LLM feedback (Google's PAT) provided by NeurIPS, which caught a minor issue in part of our proof of Theorem 2, and proposed an alternative solution for that part. Although we fixed the proof differently, we have also included the fix proposed by the LLM in Appendix A.2.1 of Appendix A.2.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.