BEYOND SCALAR LEADERBOARDS: ADAPTIVE SAMPLING FOR PARETO FRONTIER IDENTIFICATION IN MULTI-OBJECTIVE LLM ARENAS

Anonymous authors
Paper under double-blind review

ABSTRACT

The evaluation of large language models (LLMs) is inherently multi-objective, as different models may offer different tradeoffs across performance dimensions such as math, code, and writing. Because such evaluations are often based on costly and time-consuming human feedback, it is important to draw reliable conclusions using as few comparisons as possible. In this paper, we present *Vector-Borda Empirical-Gap Elimination (VB-EGE)*, a fixed-confidence adaptive sampling algorithm for efficiently identifying models on the Pareto frontier in arena-style evaluation, where each comparison evaluates a pair of models on a prompt with respect to a selected performance dimension. VB-EGE progressively eliminates models from an active set once sufficient statistical evidence has been accumulated to certify their Pareto status. Under a multi-objective Bradley–Terry model, we show that VB-EGE returns the correct Pareto frontier with high probability and derive an instance-dependent upper bound on its sample complexity. We further establish a worst-case lower bound showing that VB-EGE is minimax-optimal up to logarithmic factors. Simulations and experiments on three real-world Arena datasets (Arena, SearchArena, and VisionArena) demonstrate VB-EGE's empirical effectiveness.

1 INTRODUCTION

Pairwise human-preference judgments have become a standard tool to evaluate large language models (LLMs). In arena-style evaluation, a user submits a prompt, views two responses with the model identities hidden, and selects the preferred response (Zheng et al., 2023; Chiang et al., 2024). Given these pairwise comparisons, a common approach is to fit a statistical model to estimate a latent score for each model, and use the estimated scores to rank the models. But model quality it not one-dimensional: a model preferred for code generation need not be preferred for mathematical reasoning, creative writing, search, or vision (Liang et al., 2023; Frick et al., 2025). How should models be compared without collapsing these tradeoffs into a scalar score?

A natural target is the strict Pareto frontier: a model lies outside the frontier only if another model has a strictly higher latent score on every objective. This criterion excludes uniformly inferior models, while retaining models with different cross-objective strengths. Identifying this frontier presents a statistical challenge. Human-preference comparisons are costly to collect, and multi-objective evaluation requires sufficient samples across different objectives. We therefore develop an adaptive sampling procedure that identifies the strict Pareto frontier at a prescribed confidence level using as few comparisons as possible. As a preview, Fig. 1 presents an application of our method to comparisons from Arena-Human-Preference-140K (Chiang et al., 2024), collected through July 2025. Treating code, math, and instruction following as three objectives, the method identifies three models on the strict Pareto frontier.

1.1 PROBLEM FORMULATION

We formalize this task as a fixed-confidence sequential identification problem. Consider $K \geq 2$ candidate models evaluated on d objectives. Suppose that each model $i \in [K] := \{1, \ldots, K\}$ is

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

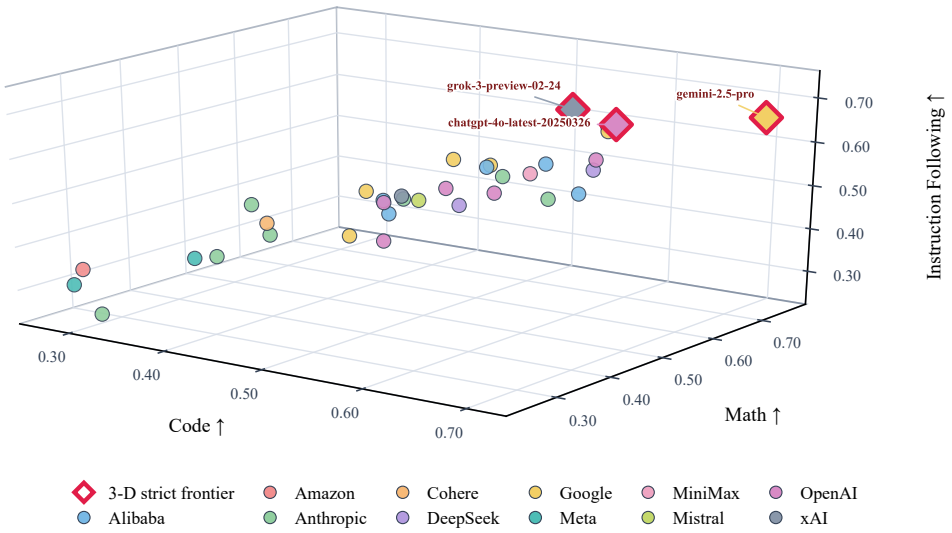


Figure 1: Three-dimensional comparison along the Code, Math, and Instruction-Following objectives from the Arena-Human-Preference-140K dataset (Chiang et al., 2024). Diamonds denote the strict Pareto frontier.

characterized by an unknown latent score vector $\theta_i = (\theta_{i,1}, \ldots, \theta_{i,d})$, with larger values indicating better performance along every coordinate.

Let $\mathcal{F}_{t-1}$ denote the observations and algorithmic randomization revealed before round t. At round t, the learner uses a possibly randomized policy based on $\mathcal{F}_{t-1}$ to select a query $Q_t = (I_t, J_t, R_t)$, where $I_t, J_t \in [K]$ are distinct models and $R_t \in [d]$ indexes the evaluation metric. For each possible query (i, j, r), let $Y_t^{(r)}(i, j) \in \{0, 1\}$ denote the potential comparison outcome at round t, with value 1 indicating that i is preferred over j on objective r. Only the outcome of the selected query is observed: $Y_t := Y_t^{(R_t)}(I_t, J_t)$. Conditional on the past history and the chosen query, the comparison outcome follows the Bradley–Terry model (Bradley & Terry, 1952):

$$\mathbb{P}_\theta\left(Y_t = 1 \,|\, \mathcal{F}_{t-1}, Q_t = (i, j, r)\right) = p_{ij}^{(r)}(\theta) := \sigma(\theta_{i,r} - \theta_{j,r}), \qquad \text{where } \sigma(x) := \frac{1}{1 + e^{-x}}. \quad (1)$$

Consequently, $p_{ji}^{(r)} = 1 - p_{ij}^{(r)}$; for convenience, we set $p_{ii}^{(r)} = 1/2$. Each query yields a single objective-specific binary label and incurs unit cost. Following Kone et al. (2024, Def. 1), we define the *strict Pareto frontier*.

Definition 1.1 (Strict Pareto frontier)**.** For a collection of vectors $x = (x_1, \ldots, x_K) \in (\mathbb{R}^d)^K$, we say that j strictly dominates i and write $i \prec_x j$, if $x_{i,r} < x_{j,r}$ for all $r \in [d]$. The strict Pareto frontier of x is

$$\mathcal{P}(x) := \{i \in [K] : \nexists j \in [K] \text{ such that } i \prec_x j\}.$$

Remark 1.1. Our strict-dominance convention differs slightly from the standard Pareto convention, under which a model dominates another if it is *at least as good* in every coordinate and strictly better in at least one coordinate (Auer et al., 2016). This convention also has a natural decision-making interpretation: if a user cares only about an objective on which two models are tied, neither model should be discarded on performance grounds; token cost, availability, or other application-specific considerations may instead determine the final choice. From a technical perspective, non-strict dominance requires additional assumptions to handle ties on individual coordinates. Our finite-sample termination guarantee requires all classification gaps defined in Def. 2.2 to be positive, which may fail at coordinate-wise ties.

Accordingly, our goal is to design an algorithm that adaptively chooses the query (I_t, J_t, R_t), stops at a stopping time τ (with $\tau = \infty$ if it never stops), and outputs an estimator $\hat{\mathcal{P}}$ of the strict Pareto frontier $\mathcal{P}^\star := \mathcal{P}(\theta)$. For a prescribed error tolerance $\delta \in (0, 1)$, we require the algorithm to be δ-correct, i.e., to terminate and return the true frontier with probability at least $1 - \delta$. The requirement is formalized below.

Definition 1.2 (δ-correctness)**.** Let $\Theta \subseteq (\mathbb{R}^d)^K$ be an instance class. For $\delta \in (0,1)$, an algorithm is δ-correct over Θ if, for every $\theta \in \Theta$,

$$\mathbb{P}_\theta\big(\tau < \infty,\ \widehat{\mathcal{P}} = \mathcal{P}(\theta)\big) \geq 1 - \delta.$$

1.2 Our contributions

Our contributions are threefold. First, we formulate fixed-confidence Pareto-set identification under objective-specific Bradley–Terry feedback and introduce *Vector-Borda Empirical-Gap Elimination (VB-EGE)*, an adaptive sampling procedure for this problem. The key technical insight is an order-preserving reduction from the latent score vectors to vector-Borda scores. Each Borda coordinate is the average win probability of a model against all opponents on a given object, and the transformation preserves the coordinate-wise ordering of the latent scores—and therefore the strict Pareto set. Moreover, each Borda coordinate can be estimated directly from Bernoulli outcomes by comparing a model with a uniformly sampled opponent. This problem therefore reduces to to Pareto-set identification over Kd-dimensional vectors of Bernoulli means. VB-EGE builds on this reduction to adaptively concentrates comparisons on models whose Pareto status remains difficult to certify.

Second, we prove that VB-EGE is δ-correct and establish an instance-dependent upper bound on its sample (comparison) complexity. We further establish a worst-case lower bound over a gap-separated instance class, showing that VB-EGE is minimax-optimal with respect to the number of models, the number of objectives, and the inverse squared classification gap, up to logarithmic factors.

Finally, we evaluate VB-EGE on synthetic Bradley–Terry instances and logged comparisons from three real-world Arena datasets: Arena, SearchArena, and VisionArena. The synthetic experiments confirm the predicted scaling behavior and demonstrate the sample efficiency of VB-EGE. Arena experiments also demonstrate substantial gains in sample efficiency over uniform baselines.

1.3 Related work

We review three related literatures, arena-based LLM evaluation, adaptive pairwise comparison, and Pareto set identification, and provide additional discussion on related work in App. A.

Arena-based LLM evaluation. Chatbot Arena aggregates human pairwise preferences into scalar model leaderboards (Zheng et al., 2023; Chiang et al., 2024). Subsequent work has improved their statistical stability by modeling ties and score dependence (Ameli et al., 2025), replacing iterative Elo updates with maximum-likelihood estimation and accounting for annotator reliability (Liu et al., 2025), and developing nonparametric ranking and uncertainty quantification (Frauen et al., 2026). Adaptive model-pair selection improves comparison efficiency (Liu et al., 2026), while prompt- or task-specific rankings address heterogeneity (Frick et al., 2025; Li et al., 2026). These methods still produce scalar rankings, either overall or conditional on prompts or tasks. We instead identify the set of models that remain nondominated across multiple objectives.

Pure exploration from pairwise comparisons. Pairwise-comparison methods study winner and ranking identification without directly observing item utilities. Random-opponent comparisons lead to Borda scores and have been used for Borda-winner identification and active ranking (Jamieson et al., 2015; Heckel et al., 2019); simple Borda-type procedures can also be robust beyond parametric Bradley–Terry models (Shah & Wainwright, 2018). More recently, Goldberger & Rudi (2026) developed an asymptotically optimal fixed-confidence method for top-k identification under a latent-utility comparison model.

Beyond these scalar targets, comparison-based methods have also been used for Pareto or skyline recovery. Asudeh et al. (2015) adaptively select criterion-specific crowd comparisons to identify Pareto-optimal objects, Groz et al. (2020) study skyline recovery under noisy coordinate comparisons, and Audiffren & Ralaivola (2017) identify maximal arms from dueling feedback over a partial order. Our setting instead considers objective-specific Bradley–Terry comparisons generated by latent utilities and uses an order-preserving vector-Borda reduction for Pareto-set identification.

Pareto set identification. Pareto-set identification has primarily been studied in multi-objective bandits where pulling an arm directly yields a vector-valued observation. Existing work considers approximate and relaxed fixed-confidence recovery (Auer et al., 2016; Kone et al., 2023), fixed-budget identification through Empirical-Gap Elimination (Kone et al., 2024), and exact fixed-confidence

identification with asymptotically optimal allocation (Crépon et al., 2024). Structured extensions include preference-cone and multi-output linear models (Karagözlü et al., 2024; Kone et al., 2025; Shukla & Basu, 2024; Das et al., 2025). Our elimination rule is closest to (Kone et al., 2024), but our feedback is objective-specific and binary rather than a directly observed vector.

2 VECTOR-BORDA EMPIRICAL-GAP ELIMINATION (VB-EGE)

In this section, we present our proposed sampling algorithm for identifying the Pareto frontier. The proposed algorithm, termed *Vector-Borda Empirical-Gap Elimination (VB-EGE)*, first represents the ordering of the models along each objective coordinate using the *Borda scores* (to be defined shortly), and then applies the empirical-gap elimination principle (Kone et al., 2024) to the vector of Borda scores to determine the sampling and stopping rule.

Notation. For any $x \in \mathbb{R}$, denote $x^+ = \max\{x, 0\}$. For $a, b \in \mathbb{R}$, write $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. For two nonnegative functions f and g, we write $f = O(g)$ (or $f \lesssim g$) if there exists a universal constant $C > 0$ such that $f \le Cg$; we write $f = \Omega(g)$ (or $f \gtrsim g$) if $g = O(f)$, and $f = \Theta(g)$ (or $f \asymp g$) if both $f = O(g)$ and $f = \Omega(g)$ hold. The notation $f = \tilde{O}(g)$, $f = \tilde{\Omega}(g)$, and $f = \tilde{\Theta}(g)$ is defined analogously, up to logarithmic factors. For $p, q \in [0, 1]$, define the binary relative entropy $\mathrm{kl}(p, q) := p \log \frac{p}{q} + (1 - p) \log \frac{1-p}{1-q}$, with the conventions $0 \log(0/x) := 0$ for $x \in [0, 1]$ and $x \log(x/0) := +\infty$ for $x > 0$.

2.1 THE VECTOR-BORDA REDUCTION

We first define the coordinate-wise Borda scores used by the algorithm.

Definition 2.1 (Coordinate-wise Borda scores)**.** For each model $i \in [K]$ and objective $r \in [d]$, define its Borda score as

$$b_{i,r}(\theta) := \frac{1}{K-1} \sum_{j \neq i} p_{ij}^{(r)}(\theta) = \frac{1}{K-1} \sum_{j \neq i} \sigma(\theta_{i,r} - \theta_{j,r}).$$

When the dependence on θ is clear from the context, we write $b_{i,r}$ in place of $b_{i,r}(\theta)$. The *vector-Borda representation* of model i is $b_i := (b_{i,1}, \ldots, b_{i,d}) \in [0, 1]^d$, and we write $b = (b_1, \ldots, b_K)$.

The key property of the vector-Borda transformation is that it preserves the ordering of the models along each coordinate induced by the latent scores θ. This property is formalized by the following lemma, with its proof provided in App. B.2.1.

Lemma 2.1 (Coordinate-wise order preservation)**.** *For every $i, j \in [K]$ and $r \in [d]$,*

$$\mathrm{sign}(b_{i,r} - b_{j,r}) = \mathrm{sign}(\theta_{i,r} - \theta_{j,r}). \tag{2}$$

Consequently,

$$i \prec_b j \iff i \prec_\theta j, \qquad \mathcal{P}(b) = \mathcal{P}(\theta) = \mathcal{P}^\star. \tag{3}$$

By Lemma 2.1, the latent strict Pareto frontier coincides with the Pareto frontier of the vector-valued Bernoulli means $b_1, \ldots, b_K$. Hence, the learner can operate entirely in Borda space without estimating the latent score vectors $\theta_1, \ldots, \theta_K$.

For a selected focal model i and objective r, draw $J \sim \mathrm{Unif}([K] \setminus \{i\})$ using fresh randomness, issue the query $Q_t = (i, J, r)$, and observe $Z := Y_t^{(r)}(i, J)$. Conditional on the past and the selected (i, r), averaging over J gives $Z \sim \mathrm{Bernoulli}(b_{i,r})$. This observation underlies our proposed sampling rule and is the coordinate-wise analogue of the standard random-opponent Borda reduction (Jamieson et al., 2015, Def. 1).

2.2 PARETO CLASSIFICATION GAPS

Next, we define classification gaps that measure how difficult it is to determine each model's Pareto status, while accounting for its role in identifying other models as non-Pareto. For $x = (x_1, \ldots, x_K) \in (\mathbb{R}^d)^K$ and any models i, j, define

$$m_x(i, j) := \min_{r \in [d]} (x_{j,r} - x_{i,r}), \qquad M_x(i, j) := \max_{r \in [d]} (x_{i,r} - x_{j,r}).$$

Thus, $M_x(i,j) = -m_x(i,j)$, and $i \prec_x j \iff m_x(i,j) > 0$.

Because our sampling algorithm follows a successive-elimination scheme, models are sequentially removed from an active set. Subsequent comparisons use only remaining models as focal models, while opponents are drawn uniformly from the full original set of models excluding the focal model. We therefore define the classification gaps relative to an arbitrarily given set $S \subseteq [K]$ to characterize the difficulty of identifying the Pareto frontier from the remaining models. First, the strict Pareto set restricted to S is given by

$$\mathcal{P}_S(x) := \{i \in S : \nexists j \in S \setminus \{i\} \text{ such that } m_x(i,j) > 0\}.$$

The definition of the classification gap is then given as follows.

Definition 2.2 (Restricted-set classification gaps)**.** Let $S \subseteq [K]$ with $|S| \geq 2$, and let $x_i \in \mathbb{R}^d$ for every $i \in S$. For every $i \in S$, define the raw domination gap $\Delta^\star_{i,S}(x) := \max_{j \in S\setminus\{i\}} m_x(i,j)$, and, for every $i \in S$, the auxiliary gap

$$\delta^\star_{i,S}(x) := \min_{j \in S\setminus\{i\}} \left\{ M_x(i,j) \wedge \left[M_x(j,i)^+ + \left(\Delta^\star_{j,S}(x)\right)^+ \right] \right\}.$$

The restricted-set classification gap is

$$\Delta_{i,S}(x) := \begin{cases} \Delta^\star_{i,S}(x), & i \notin \mathcal{P}_S(x), \\ \delta^\star_{i,S}(x), & i \in \mathcal{P}_S(x). \end{cases}$$

Intuitively, for $i \notin \mathcal{P}_S(x)$, $\Delta^\star_{i,S}(x)$ measures the largest domination margin certifying i is non-Pareto. For $i \in \mathcal{P}_S(x)$, $\delta^\star_{i,S}(x)$ is the smallest separation needed to certify i's Pareto status or to correctly classify another model relative to i. Thus, a larger $\Delta_{i,S}(x)$ indicates estimating model i for identifying $\mathcal{P}_S(x)$ is easier.

In the following, we omit the set subscript when $S = [K]$ and write $\mathcal{P}(x) := \mathcal{P}_{[K]}(x)$, $\Delta^\star_i(x) := \Delta^\star_{i,[K]}(x)$, $\delta^\star_i(x) := \delta^\star_{i,[K]}(x)$, $\Delta_i(x) := \Delta_{i,[K]}(x)$. For the vector-Borda and latent representations, respectively, let $\Delta^{\mathrm{B}}_i := \Delta_i(b)$, $\Delta^\theta_i := \Delta_i(\theta)$, and $\Delta^{\mathrm{B}}_{\min} := \min_{i\in[K]} \Delta^{\mathrm{B}}_i$.

2.3 Phased vector-Borda elimination

VB-EGE proceeds in phases, maintaining an active set A of models whose Pareto status remains undetermined and an accepted set B of models certified as Pareto. At each phase $m \geq 1$, for each pair $(i,r) \in A \times [d]$, the algorithm collects $n_m - n_{m-1}$ new observations by sampling

$$Z_{i,r,s} := Y^{(r)}_{t(i,r,s)}(i, J_{i,r,s}), \text{ where } J_{i,r,s} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}([K]\setminus\{i\}), \quad \forall s = n_{m-1}+1, \ldots, n_m, \quad (4)$$

where $t(i,r,s)$ denotes the global round at which the sth comparison associated with model i and coordinate r is conducted, $n_0 = 0$, and $\{n_m\}_{m\geq 1}$ is an increasing sequence of sample-size parameters specified below. Using all the observations collected through phase m, the algorithm updates the vector-Borda estimates as

$$\widehat{b}^{(m)}_{i,r} := \frac{1}{n_m} \sum_{s=1}^{n_m} Z_{i,r,s}, \quad (i,r) \in A \times [d].$$

Write $\widehat{b}^{(m)}_i = (\widehat{b}^{(m)}_{i,1}, \ldots, \widehat{b}^{(m)}_{i,d})$; all empirical Pareto sets and gaps below are evaluated using these vectors for models in the current active set A. It then updates the restricted-set Pareto frontier and the corresponding classification gap with respect to the current active set A:

$$\widehat{\mathcal{S}}_A := \mathcal{P}_A\big(\widehat{b}^{(m)}\big), \qquad \widehat{\Delta}_{i,A} := \Delta_{i,A}\big(\widehat{b}^{(m)}\big), \quad i \in A. \quad (5)$$

If all the estimated classification gaps in A are small—that is, $\max_{i\in A} \widehat{\Delta}_{i,A} \leq C_{\mathrm{thr}} r_m$, where C_{thr} is a constant and r_m is a confidence-radius parameter specified below—then the available evidence is not sufficient to certify the status of any additional model and the algorithm proceeds to the next phase. Otherwise, it selects $i_m \in \arg\max_{i\in A} \widehat{\Delta}_{i,A}$ and removes it from A. The selected model i_m is chosen anew at each iteration of the inner loop within phase m. Ties are resolved by prioritizing

empirically non-Pareto models, as specified in Alg. 1. If $i_m \in \widehat{\mathcal{S}}_A$, it is certified as Pareto and added to B; otherwise, it is rejected as non-Pareto. After each removal, if only one active model remains, the algorithm returns $B \cup A$. Otherwise, it recomputes the empirical Pareto set and classification gaps over the updated active set and repeats the removal step, using the same phase-m estimates without additional sampling. It proceeds to phase $m+1$ only when no empirical gap exceeds the threshold.

The sample sizes and removal thresholds are chosen so that, with high probability, every model added to B is truly Pareto and every rejected model is truly non-Pareto. To be specific, at each phase $m \geq 1$, we define

$$\varepsilon_m := 2^{-m}, \qquad L_m := \log \frac{C_{\log} K d m^2}{\delta}, \qquad n_m := \left\lceil \frac{C_{\mathrm{samp}} L_m}{\varepsilon_m^2} \right\rceil, \qquad r_m := \sqrt{\frac{L_m}{2n_m}}, \tag{6}$$

where $C = (C_{\mathrm{samp}}, C_{\mathrm{thr}}, C_{\log})$ are hyperparameters controlling the sampling rate, removal threshold, and confidence factor of the algorithm. Unless stated otherwise, we use the default configuration

$$C = (C_{\mathrm{samp}}, C_{\mathrm{thr}}, C_{\log}) = (2, 4, 4). \tag{7}$$

The complete procedure is summarized in Alg. 1.

Remark 2.1 (Sampling and tie-breaking conventions). VB-EGE samples only active models as focal models, but draws each opponent uniformly from the full original set of models $[K] \setminus \{i\}$, including previously removed models. Thus the mean of each focal observation remains the fixed global quantity $b_{i,r}$; restricting opponents to the active set would instead produce an active-set-dependent Borda coordinate. When multiple models maximize $\widehat{\Delta}_{i,A}$, VB-EGE prioritizes a maximizer outside the empirical Pareto set $\widehat{\mathcal{S}}_A$, whenever one exists; otherwise, it selects an arbitrary maximizer. On the uniform confidence event in Lemma B.6, this tie-breaking rule ensures that each active non-Pareto model i retains an active Pareto model attaining its domination gap Δ_i^{B}.

Algorithm 1 Phased Vector-Borda Empirical-Gap Elimination VB-EGE$(\delta; C)$

Require: number of candidate models $K \geq 2$, number of objectives $d \geq 1$, error tolerance level $\delta \in (0,1)$, and sampling hyperparameters $C = (C_{\mathrm{samp}}, C_{\mathrm{thr}}, C_{\log}) \in (0,\infty)^2 \times [1,\infty)$.

Initialization: $A \leftarrow [K]$, $B \leftarrow \varnothing$, $n_0 \leftarrow 0$.
for $m = 1, 2, \ldots$ **do**
  Set $\varepsilon_m, L_m, n_m, r_m$ according to (6).
  for $(i, r) \in A \times [d]$ **do**
    Draw $\{Z_{i,r,s}\}_{s=n_{m-1}+1}^{n_m}$ according to (4), and compute $\widehat{b}_{i,r}^{(m)} \leftarrow \dfrac{1}{n_m} \displaystyle\sum_{s=1}^{n_m} Z_{i,r,s}$.
  end for
  Compute $\widehat{\mathcal{S}}_A$ and $\{\widehat{\Delta}_{i,A}\}_{i \in A}$ with $\{\widehat{b}_i^{(m)} : i \in A\}$ according to (5).
  while $|A| \geq 2$ and $\max_{i \in A} \widehat{\Delta}_{i,A} > C_{\mathrm{thr}} r_m$ **do**
    $G \leftarrow \arg\max_{i \in A} \widehat{\Delta}_{i,A}$.
    Choose $i_m \in G \setminus \widehat{\mathcal{S}}_A$ if $G \setminus \widehat{\mathcal{S}}_A \neq \varnothing$; otherwise choose $i_m \in G$.
    if $i_m \in \widehat{\mathcal{S}}_A$ **then**
      $B \leftarrow B \cup \{i_m\}$
    end if
    $A \leftarrow A \setminus \{i_m\}$.
    if $|A| = 1$ **then return** $B \cup A$.
    Recompute $\widehat{\mathcal{S}}_A$ and $\{\widehat{\Delta}_{i,A}\}_{i \in A}$ over the updated A with $\{\widehat{b}_i^{(m)} : i \in A\}$.
  end while
end for

3 Theoretical Guarantees

This section presents the theoretical guarantees for VB-EGE. For an algorithm A, let τ_{A} denote the total number of queries issued before it stops, with $\tau_{\mathsf{A}} = \infty$ if it does not stop. Recall that each query compares two models with respect to a single objective and returns a binary outcome.

Since τ_{A} is random, we characterize its sample complexity through high-probability stopping-time quantiles. Specifically, for $\alpha \in (0,1)$, define $Q_{\theta,\alpha}(\mathsf{A}) := \inf\{T \in \mathbb{N}_0 : \mathbb{P}_\theta(\tau_{\mathsf{A}} \le T) \ge 1-\alpha\}$, with the convention $\inf \varnothing := \infty$. When the algorithm is clear from the context, we write $\tau = \tau_{\mathsf{A}}$.

For $B, \Delta > 0$, define the bounded-range class and its gap-separated subclass by

$$\Theta_{B,d} := \left\{\theta \in (\mathbb{R}^d)^K : \max_{r\in[d]} \max_{i,j\in[K]} |\theta_{i,r} - \theta_{j,r}| \le 2B\right\}, \Theta_{B,d}(\Delta) := \left\{\theta \in \Theta_{B,d} : \min_{i\in[K]} \Delta_i^\theta \ge \Delta\right\}.$$

We also write $\kappa_B := \inf_{|x|\le 2B} \sigma'(x) = \sigma(2B)\big(1-\sigma(2B)\big)$. In what follows, Sec. 3.1 establishes the δ-correctness of VB-EGE and an instance-dependent upper bound on its sample complexity over the instance class $\Theta_{B,d}$. Sec. 3.2 provides a worst-case lower bound over $\Theta_{B,d}(\Delta)$. Together, these results show that VB-EGE is minimax-optimal over $\Theta_{B,d}(\Delta)$ up to logarithmic factors.

3.1 Upper bound

The following theorem establishes the δ-correctness of our proposed algorithm and an upper bound on the quantile of its stopping time.

Theorem 3.1 (Quantile upper bound). *Fix $B > 0$ and $\theta \in \Theta_{B,d}$ satisfying $\min_{i\in[K]} \Delta_i^\theta > 0$. Define for each $i \in [K]$ $M_i^\theta := 1 + \left\lceil \log_2 \frac{4}{\kappa_B \Delta_i^\theta} \right\rceil$, $L_i^\theta(\delta) := \log \frac{4Kd(M_i^\theta)^2}{\delta}$. Then VB-EGE (Alg. 1) implemented with $C = (2,4,4)$ satisfies*

$$\mathbb{P}_\theta\left(\widehat{\mathcal{P}} = \mathcal{P}(\theta), \tau \le dK + \frac{128d}{\kappa_B^2} \sum_{i=1}^K \frac{L_i^\theta(\delta)}{(\Delta_i^\theta)^2}\right) \ge 1-\delta.$$

Consequently, VB-EGE is δ-correct on the positive-gap subclass of $\Theta_{B,d}$, and

$$Q_{\theta,\delta}(\text{VB-EGE}) \le dK + \frac{128d}{\kappa_B^2} \sum_{i=1}^K \frac{L_i^\theta(\delta)}{(\Delta_i^\theta)^2} = \widetilde{O}\left(\frac{d}{\kappa_B^2} \sum_{i=1}^K (\Delta_i^\theta)^{-2}\right).$$

If B is bounded by an absolute constant, then κ_B is bounded below by a positive absolute constant.

The proof first establishes exact recovery and finite stopping in Borda space, then converts the bound to latent-score gaps; see App. B.2.2. The factor $\kappa_B^{-2} = \Theta(e^{4B})$ comes from a worst-case gap conversion and may be conservative.

3.2 Lower bound

We now turn to the lower bound on the sample complexity of any δ-correct algorithm. For the gap-separated instance class $\Theta_{B,d}(\Delta)$, define the minimax $(1-\delta)$-quantile sample complexity as

$$\mathbf{Q}_{B,d}(\Delta,\delta) := \inf_{\mathsf{A}:\ \delta\text{-correct on } \Theta_{B,d}(\Delta)} \sup_{\theta\in\Theta_{B,d}(\Delta)} Q_{\theta,\delta}(\mathsf{A}).$$

Thus, $\mathbf{Q}_{B,d}(\Delta,\delta)$ is the smallest worst-case $(1-\delta)$-stopping-time quantile achievable by any δ-correct algorithm on $\Theta_{B,d}(\Delta)$. The following theorem gives a lower bound on $\mathbf{Q}_{B,d}(\Delta,\delta)$.

Theorem 3.2 (Gap-separated minimax lower bound). *Fix $B > 0$, $0 < \Delta \le \min\{1, B\}$, and $0 < \delta < 1/3$. Then $\mathbf{Q}_{B,d}(\Delta,\delta) \ge \frac{d(K-1)}{\Delta^2} \operatorname{kl}(1-2\delta,\delta)$.*

The proof of Thm. 3.2 is deferred to App. B.2.3. Here, we sketch the high-level idea: the proof considers the hard instance, where for some $i \in [K]$,

$$\theta_i = (0,\dots,0), \quad \theta_j = (-\Delta,\dots,-\Delta), \quad \forall j \in [K]\backslash\{i\}. \tag{8}$$

Raising any model other than i to $+\Delta$ on a single objective makes it Pareto, while preserving the gap and range constraints. A correct algorithm must rule out all $d(K-1)$ such possibilities.

Finally, applying Thm. 3.1 with $\Delta_i^\theta \ge \Delta$ for every i and combining it with Thm. 3.2 yields the following minimax guarantee for fixed B, up to logarithmic factors.

Corollary 3.3 (Gap-separated minimax optimality)**.** *Under the same assumptions as in Thm. 3.2,*

$$\mathbf{Q}_{B,d}(\Delta,\delta) \le \sup_{\theta\in\Theta_{B,d}(\Delta)} Q_{\theta,\delta}(\text{VB-EGE}) \le dK + \frac{128dK}{\kappa_B^2\Delta^2}\log\frac{4Kd(\lceil\log_2\frac{8}{\kappa_B\Delta}\rceil)^2}{\delta}.$$

Consequently, for any fixed $B > 0$, $\mathbf{Q}_{B,d}(\Delta,\delta) = \widetilde{\Theta}\big(\frac{dK}{\Delta^2}\log\frac{1}{\delta}\big)$ as $\delta \downarrow 0$, where $\widetilde{\Theta}$ suppresses logarithmic factors in K, d, and Δ^{-1}. In particular, VB-EGE attains the minimax rate up to these logarithmic factors.

4 Experiments

This section evaluates the scaling and sample efficiency of VB-EGE on synthetic data, as well as its sample efficiency on logged Arena comparisons.

4.1 Synthetic experiments

In the scaling experiments, we generate data from the BT model corresponding to the hard instance in (8) and vary the values of K, d, Δ, δ. Separately, we evaluate the methods on eight synthetic benchmarks generated from four families, which we call *Convex, Arena, Witness*, and *Two-group*. Details of the data-generating process in both experiments are provided in App. C.1.

We compare VB-EGE with three baselines using uniform sampling:
(1) *Focal Borda*. At each phase, the algorithm collects new observations for *every* model-objective pair $(i, r) \in [K] \times [d]$ and performs no elimination; otherwise, it uses the same sampling schedule $\{n_m\}_{m\ge1}$, Borda estimates, and certification rule as VB-EGE.

(2) *BT-MLE*. The algorithm uniformly samples from all $dK(K-1)/2$ pair–objective tuples (i, j, r), where $1 \le i < j \le K$ and $r \in [d]$. At the end of each phase, it fits a separate ridge-regularized BT model for each objective to estimate the latent scores. As a plug-in heuristic, it substitutes these estimates into the classification-gap function and compares the resulting gaps with the same threshold as VB-EGE to determine the estimated Pareto set and whether to stop.

(3) *Borda plug-in*. The algorithm uses the same sampling rule as BT-MLE but estimates the Borda scores directly using smoothed empirical win rates, whose exact definition is given in App. C.2. The stopping and Pareto certification rule follows from VB-EGE.

All methods use $\delta = 0.05$ and $C = (2, 4, 4)$, except when we evaluate the scaling with respect to δ. Other implementation details can be found in App. C.2. Each experiment is repeated over 500 independent trials, and we report the mean stopping time, which counts objective-specific counts, and proportion of trials in which the Pareto set is incorrectly classified.

Results. Across all synthetic settings, we observe no errors in identifying the Pareto set. In Fig. 2a, the empirical log–log slopes of the mean stopping time with respect to K and d are both 1.09 for VB-EGE, consistent with the scaling predicted by our theory. The mean stopping time also decreases as Δ increases and is nearly linear in $\log(1/\delta)$, with $R^2 \approx 0.99997$. In these experiments, VB-EGE and Focal Borda have nearly identical performance, which is as expected because all dominated models are equally difficult to classify, leaving little to gain from adaptive elimination.

Fig. 2b shows the results from the eight heterogeneous benchmarks. VB-EGE achieves the lowest mean stopping cost in all settings. For each baseline and trial, we compute the ratio between the baseline's stopping time to that of VB-EGE and then average these ratios over trials. Across the eight settings, the resulting mean ratio ranges from 1.26 to 17.9 for Focal Borda, from 1.80 to 38.7 for BT-MLE, and from 29.8 to 649 for Borda plug-in. App. C presents additional results and diagnostics.

4.2 Realistic Arena data experiments

We consider Arena-Human-Preference-140K (Chiang et al., 2024), Search-Arena-24K (Miroyan et al., 2026), and VisionArena-Battle (Chou et al., 2025), treating the released task tags, search intents, and vision categories as objectives, respectively. Within each dataset, we retain only decisive comparisons—those in which either model is labeled the winner—and exclude the ties and "both bad"

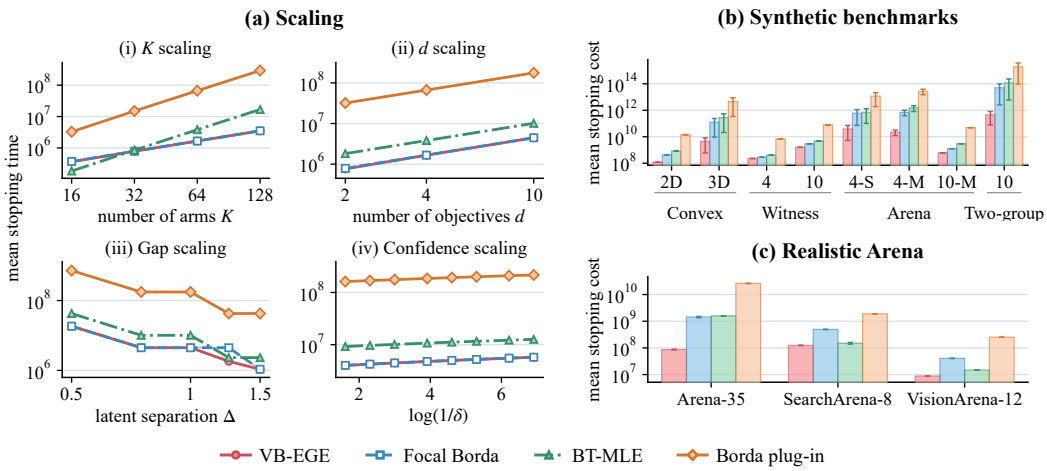


Figure 2: Mean stopping time (mean $\pm$ one standard error) on logarithmic vertical axes. (a) Scaling in K, d, Δ, and $\log(1/\delta)$. (b) Eight heterogeneous synthetic Pareto benchmark settings; S and M denote small and medium. (c) Three realistic Arena datasets, with stopping times measured in comparisons.

outcomes. We then select the largest subset of models such that every pair has at least one retained comparison for every objective. Additional preprocessing details are provided in App. D.1.

For each model and objective, we compute a Borda score by averaging its logged win rates against all other retained models. We take the strict Pareto frontier of the resulting Borda vectors as the empirical ground truth. At each sampling step, after selecting a model pair and an objective, one binary outcome is drawn uniformly *with replacement* from the corresponding logged outcomes. We evaluate VB-EGE and the three baseline methods described in Section 4.1, using $\delta = 0.05$, $C = (2, 4, 4)$. Each experiment is repeated over 100 independent trials.

Results. Across all experiments, every method stopped and correctly identified the empirical Pareto frontier in every trial. Figure 2c compares the mean stopping time, where VB-EGE achieves the lowest value on all three datasets. The mean baseline-to-VB-EGE stopping-time ratio ranges from 3.96 to 16.54 for Focal Borda, from 1.20 to18.27 for BT-MLE, and from 15.20 to 301.36 for Borda plug-in. Additional results and discussion are provided in App. D.2 and App. D.4.

5 Limitations and Future Work

This paper introduced a fixed-confidence adaptive sampling framework for identifying Pareto-optimal models from pairwise comparison. Several directions remain for future work. First, although our guarantees are stated under the BT model, the order-preservation property of the vector-Borda reduction requires only that the comparison probability to be strictly increasing in the latent score difference. The reduction therefore applies more broadly to monotone comparison models. Second, our theory considers only decisive comparisons and does not accommodate the `tie` and `both_bad` labels in Chatbot Arena (Chiang et al., 2024). App. D.4 examines sensitivity to alternative treatments of these non-decisive labels; a rigorous theoretical treatment is left for future work. Third, the Arena experiments use several forms of objective labeling, including mutually exclusive groups, primary-intent labels, and overlapping categories. An important open question is how sensitive the identified frontiers are to these choices. One extension is to allow soft or learned assignments, where each comparison context is associated with a mixture of capabilities.

Beyond VB-EGE, there are also several algorithmic directions worth investigating. For example, under the BT model, one could estimate the latent scores and guide sampling using likelihood-based methods. App. E explores a Pareto-specific heuristic of this kind, leaving a rigorous theoretical analysis for future work. Another promising direction is graph-aware sampling that exploits nonuniform historical support when selecting opponents and objectives. Such methods would be particularly relevant for larger, evolving arenas in which comparison coverage is uneven and new models are introduced over time.

REPRODUCIBILITY

Code for reproducing the results in this paper is available at https://anonymous.4open.science/r/Beyond_Scalar_Leaderboards-F73D. Dataset links, pinned revisions, and preprocessing details are provided in App. D.1.

AI USE STATEMENT

In this work, we used generative AI tools (including ChatGPT 5.6 Sol, ChatGPT 6 Astra, etc.) for organizing the paper outline, polishing the writing, and structuring the appendices; searching for relevant literature; implementing synthetic-data generation and sampling from logged Arena comparisons; and, after the upper-bound proof had been established, searching for and assisting with the derivation of a matching lower bound. We have not used generative AI tools for designing the conceptual frameworks, proposing or refining hypotheses or interpreting results. We have reviewed all technical claims, proofs, experimental designs, implementations, and reported results. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Siavash Ameli, Siyuan Zhuang, Ion Stoica, and Michael W. Mahoney. A statistical framework for ranking LLM-based chatbots. In *International Conference on Learning Representations*, 2025.

Abolfazl Asudeh, Gensheng Zhang, Naeemul Hassan, Chengkai Li, and Gergely V. Zaruba. Crowdsourcing Pareto-optimal object finding by pairwise comparisons. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 753–762, 2015.

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *Proceedings of the 23rd Conference on Learning Theory*, pp. 41–53, 2010.

Julien Audiffren and Liva Ralaivola. Bandits dueling on partially ordered sets. In *Advances in Neural Information Processing Systems*, volume 30, pp. 2130–2139. Curran Associates, Inc., 2017.

Peter Auer, Chao-Kai Chiang, Ronald Ortner, and Madalina Drugan. Pareto front identification from stochastic bandit feedback. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pp. 939–947. PMLR, 2016.

Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3–4):324–345, 1952. doi: 10.1093/biomet/39.3-4.324.

Sébastien Bubeck, Tengyao Wang, and Nitin Viswanathan. Multiple identifications in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 258–265. PMLR, 2013.

Emil Carlsson, Debabrota Basu, Fredrik D. Johansson, and Devdatt Dubhashi. Pure exploration in bandits with linear constraints. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 334–342. PMLR, 2024.

Shouyuan Chen, Tian Lin, Irwin King, Michael R. Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, volume 27, pp. 379–387. Curran Associates, Inc., 2014.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. Chatbot Arena: An open platform for evaluating LLMs by human preference. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 8359–8388. PMLR, 2024.

Brian M. Cho, Dominik Meier, Kyra Gan, and Nathan Kallus. Reward maximization for pure exploration: Minimax optimal good arm identification for nonparametric multi-armed bandits. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 2251–2259. PMLR, 2025.

Christopher Chou, Lisa Dunlap, Koki Mashita, Krishna Mandal, Trevor Darrell, Ion Stoica, Joseph E. Gonzalez, and Wei-Lin Chiang. VisionArena: 230k real world user-VLM conversations with preference labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3877–3887, 2025.

Élise Crépon, Aurélien Garivier, and Wouter M. Koolen. Sequential learning of the Pareto front for multi-objective bandits. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3583–3591. PMLR, 2024.

Udvas Das, Apurv Shukla, and Debabrota Basu. FraPPE: Fast and efficient preference-based pure exploration. In *Advances in Neural Information Processing Systems*, volume 38, pp. 10190–10236. Curran Associates, Inc., 2025. doi: 10.52202/085713-0312.

Roger R. Davidson. On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329):317–328, 1970. doi: 10.1080/01621459.1970.10481082.

Rémy Degenne and Wouter M. Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems*, volume 32, pp. 14524–14533. Curran Associates, Inc., 2019.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(39):1079–1105, 2006.

Dennis Frauen, Athiya Deviyani, Mihaela van der Schaar, and Stefan Feuerriegel. Nonparametric LLM evaluation from preference data. *arXiv preprint arXiv:2601.21816*, 2026.

Evan Frick, Connor Chen, Joseph Tennyson, Tianle Li, Wei-Lin Chiang, Anastasios Nikolas Angelopoulos, and Ion Stoica. Prompt-to-Leaderboard: Prompt-adaptive LLM evaluations. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 17672–17689. PMLR, 2025.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pp. 998–1027. PMLR, 2016.

Motti Goldberger and Nils Rudi. Optimal top-k identification from pairwise comparisons. *arXiv preprint arXiv:2607.08979*, 2026.

Benoît Groz, Frederik Mallmann-Trenn, Claire Mathieu, and Victor Verdugo. Skyline computation with noisy comparisons. In *Combinatorial Algorithms*, volume 12126 of *Lecture Notes in Computer Science*, pp. 289–303. Springer, 2020. doi: 10.1007/978-3-030-48966-3_22.

Reinhard Heckel, Nihar B. Shah, Kannan Ramchandran, and Martin J. Wainwright. Active ranking from pairwise comparisons and when parametric assumptions do not help. *The Annals of Statistics*, 47(6):3099–3126, 2019. doi: 10.1214/18-AOS1772.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'UCB: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pp. 423–439. PMLR, 2014.

Kevin Jamieson, Sumeet Katariya, Atul Deshpande, and Robert Nowak. Sparse dueling bandits. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pp. 416–424. PMLR, 2015.

Marc Jourdan, Rémy Degenne, Dorian Baudry, Rianne de Heide, and Emilie Kaufmann. Top two algorithms revisited. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26791–26803. Curran Associates, Inc., 2022. doi: 10.52202/068431-1943.

Efe Mert Karagözlü, Yaşar Cahit Yıldırım, Çağın Ararat, and Cem Tekin. Learning the Pareto set under incomplete preferences: Pure exploration in vector bandits. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3070–3078. PMLR, 2024.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 1238–1246. PMLR, 2013.

Julian Katz-Samuels and Clay Scott. Feasible arm identification. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2535–2543. PMLR, 2018.

Emilie Kaufmann and Shivaram Kalyanakrishnan. Information complexity in bandit subset selection. In *Proceedings of the 26th Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pp. 228–251. PMLR, 2013.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17(1):1–42, 2016.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Adaptive algorithms for relaxed Pareto set identification. In *Advances in Neural Information Processing Systems*, volume 36, pp. 35190–35201. Curran Associates, Inc., 2023. doi: 10.52202/075280-1528.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Bandit Pareto set identification: The fixed budget setting. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2548–2556. PMLR, 2024.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Bandit Pareto set identification in a multi-output linear model. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 1189–1197. PMLR, 2025.

Jiachun Li, David Simchi-Levi, and Will Wei Sun. Low rank for rank: Uncertainty-aware task-specific LLM ranking under sparse pairwise comparisons. *arXiv preprint arXiv:2605.29395*, 2026.

Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, Benjamin Newman, Binhang Yuan, Bobby Yan, Ce Zhang, Christian Cosgrove, Christopher D. Manning, Christopher Ré, Diana Acosta-Navas, Drew A. Hudson, Eric Zelikman, Esin Durmus, Faisal Ladhak, Frieda Rong, Hongyu Ren, Huaxiu Yao, Jue Wang, Keshav Santhanam, Laurel Orr, Lucia Zheng, Mert Yuksekgonul, Mirac Suzgun, Nathan Kim, Neel Guha, Niladri Chatterji, Omar Khattab, Peter Henderson, Qian Huang, Ryan Chi, Sang Michael Xie, Shibani Santurkar, Surya Ganguli, Tatsunori Hashimoto, Thomas Icard, Tianyi Zhang, Vishrav Chaudhary, William Wang, Xuechen Li, Yifan Mai, Yuhui Zhang, and Yuta Koreeda. Holistic evaluation of language models. *Transactions on Machine Learning Research*, 2023.

Zirui Liu, Jiatong Li, Yan Zhuang, Qi Liu, Shuanghong Shen, Jie Ouyang, Mingyue Cheng, and Shijin Wang. am-ELO: A stable framework for arena-based LLM evaluation. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 38857–38868. PMLR, 2025.

Zirui Liu, Xianquan Wang, Yan Zhuang, Jiatong Li, Qi Liu, Shuanghong Shen, Mingyue Cheng, and Shijin Wang. Fewer battles, more gain: An information-efficient framework for arena-based LLM evaluation. In *International Conference on Learning Representations*, 2026.

Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 1690–1698. PMLR, 2016.

Shie Mannor and John N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5:623–648, 2004.

Mihran Miroyan, Tsung-Han Wu, Logan King, Tianle Li, Jiayi Pan, Xinyan Hu, Wei-Lin Chiang, Anastasios N. Angelopoulos, Trevor Darrell, Narges Norouzi, and Joseph E. Gonzalez. Search Arena: Analyzing search-augmented LLMs. In *International Conference on Learning Representations*, 2026.

Shintaro Nakamura and Masashi Sugiyama. Fixed-budget real-valued combinatorial pure exploration of multi-armed bandit. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1225–1233. PMLR, 2024.

Daniel Russo. Simple Bayesian algorithms for best-arm identification. *Operations Research*, 68(6): 1625–1647, 2020. doi: 10.1287/opre.2019.1911.

Nihar B. Shah and Martin J. Wainwright. Simple, robust and optimal ranking from pairwise comparisons. *Journal of Machine Learning Research*, 18(199):1–38, 2018.

Apurv Shukla and Debabrota Basu. Preference-based pure exploration. In *Advances in Neural Information Processing Systems*, volume 37, pp. 17313–17347. Curran Associates, Inc., 2024. doi: 10.52202/079017-0550.

Andrea Tirinzoni and Rémy Degenne. On elimination strategies for bandit fixed-confidence identification. In *Advances in Neural Information Processing Systems*, volume 35, pp. 18586–18598. Curran Associates, Inc., 2022. doi: 10.52202/068431-1350.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. In *Advances in Neural Information Processing Systems*, volume 36, pp. 46595–46623. Curran Associates, Inc., 2023. doi: 10.52202/075280-2020.

APPENDIX CONTENTS

A Additional related work **15**

B Technical proofs **15**

B.1 Supporting results 15

B.2 Proofs of main results 24

C Additional details and results for synthetic experiments **28**

C.1 Details for the data-generating process 28

C.2 Implementation details for the baseline models 29

C.3 Additional results from the synthetic benchmark settings 29

C.4 Hyperparameter sensitivity analysis 30

C.5 Detailed fixed-confidence scaling results 31

C.6 Confidence scaling and tail quantiles 31

C.7 Pareto-set size sensitivity analysis 32

C.8 Sampling-mechanism diagnostics 33

D Additional real-data experiments details and results **34**

D.1 Datasets, preprocessing, and retained models 34

D.2 Full Pareto-set results and method comparisons 36

D.3 Sampling from logged Arena comparisons 39

D.4 Sensitivity to `tie` and `both_bad` handling 43

E Additional comparison with Track-and-Stop extensions **50**

E.1 Implementation used in the reported experiments 50

E.2 Comparison with the exact stopping statistic 51

E.3 Provisional screened results 52

A Additional related work

Best arm identification Best-arm identification is a canonical pure-exploration problem in stochastic bandits, where samples are collected to identify an optimal arm rather than to maximize cumulative reward. Early work established PAC sample-complexity bounds and confidence-based elimination and stopping procedures for identifying a near-optimal arm (Mannor & Tsitsiklis, 2004; Even-Dar et al., 2006). Subsequent work developed sharper procedures for best-arm identification under fixed-budget and fixed-confidence formulations, including Successive Rejects (Audibert et al., 2010), Sequential Halving (Karnin et al., 2013), and lil'UCB (Jamieson et al., 2014). Kaufmann et al. (2016) studied the instance-dependent complexity of the two dominant fixed-budget and fixed-confidence formulations, while Garivier & Kaufmann (2016) derived a tight fixed-confidence lower bound and introduced the asymptotically optimal Track-and-Stop strategy. More recent work has developed alternative adaptive-allocation principles, notably Top-Two methods (Russo, 2020; Jourdan et al., 2022). Elimination strategies can further reduce the computational cost of fully adaptive fixed-confidence identification while preserving sample-complexity guarantees (Tirinzoni & Degenne, 2022).

Set-valued and structured identification. Beyond best-arm identification, bandit identification has been studied for richer terminal targets. Top-m and subset-selection problems seek several of the highest-mean arms (Bubeck et al., 2013; Kaufmann & Kalyanakrishnan, 2013), while thresholding and good-arm identification aim to determine which arms satisfy a prescribed performance criterion (Locatelli et al., 2016; Cho et al., 2025). Combinatorial pure exploration generalizes the answer space to structured decision classes, including fixed-size subsets, matchings, spanning trees, and paths (Chen et al., 2014), with recent work developing fixed-budget methods for more general real-valued combinatorial action classes (Nakamura & Sugiyama, 2024). Related extensions consider vector-valued observations and constrained decision spaces: Katz-Samuels & Scott (2018) study feasible-arm identification with multivariate means, while Carlsson et al. (2024) consider fixed-confidence identification when the optimal policy is subject to linear constraints and may mix several arms. More generally, Degenne & Koolen (2019) develop a fixed-confidence framework for pure-exploration problems that may admit multiple correct answers.

B Technical proofs

B.1 Supporting results

B.1.1 Properties of the classification gaps

Lemma B.1 (Basic inequalities for pairwise margins)**.** *Let* $x = (x_1, \dots, x_K), \quad y = (y_1, \dots, y_K) \in (\mathbb{R}^d)^K$. *Then, for all* $i, j, k \in [K]$,

(i) $m_x(i,k) \geq m_x(i,j) + m_x(j,k)$;

(ii) if $\max_{u\in[K]} \max_{r\in[d]} |x_{u,r} - y_{u,r}| \leq \rho$, *then*

$$|m_x(i,j) - m_y(i,j)| \leq 2\rho, \qquad |M_x(i,j) - M_y(i,j)| \leq 2\rho;$$

(iii) for all $a, b \in \mathbb{R}$, $|a^+ - b^+| \leq |a - b|$.

Proof. For part (i), coordinate-wise,

$$x_{k,r} - x_{i,r} = (x_{k,r} - x_{j,r}) + (x_{j,r} - x_{i,r}),$$

Taking the minimum over r and using $\min_r(a_r + b_r) \geq \min_r a_r + \min_r b_r$ gives

$$m_x(i,k) \geq m_x(i,j) + m_x(j,k).$$

For part (ii), every coordinate difference satisfies

$$|(x_{j,r} - x_{i,r}) - (y_{j,r} - y_{i,r})| \leq 2\rho.$$

Both the minimum and maximum maps are 1-Lipschitz with respect to the ℓ_∞ norm, which proves the two inequalities.

Part (iii) follows because the map $a \mapsto a^+$ is 1-Lipschitz. □

Lemma B.2 (Pareto witness for a non-Pareto model). *If $i \notin \mathcal{P}(x)$, then*

$$\Delta_i^\star(x) = \max_{p \in \mathcal{P}(x)} m_x(i,p) > 0.$$

In particular, there exists $i_x^\star \in \mathcal{P}(x)$ such that $m_x(i, i_x^\star) = \Delta_i^\star(x)$.

Proof. Since $i \notin \mathcal{P}(x)$, there exists a model that strictly dominates i. Hence

$$\Delta_i^\star(x) = \max_{j \in [K]\setminus\{i\}} m_x(i,j) > 0.$$

Because the model set is finite, we may choose $i_x^\star \in \arg\max_{j\in[K]\setminus\{i\}} m_x(i,j)$. Then $m_x(i,i_x^\star) = \Delta_i^\star(x) > 0$, so $i \prec_x i_x^\star$.

We claim that $i_x^\star \in \mathcal{P}(x)$. Suppose otherwise. Then there exists $p \in [K]$ such that $i_x^\star \prec_x p$, and therefore $m_x(i_x^\star, p) > 0$. By Lemma B.1,

$$m_x(i,p) \geq m_x(i,i_x^\star) + m_x(i_x^\star,p) > m_x(i,i_x^\star) = \Delta_i^\star(x).$$

This contradicts the definition of $i_x^\star$. Hence $i_x^\star \in \mathcal{P}(x)$, and because it already attains the maximum over all $j \neq i$, it also attains the maximum over $p \in \mathcal{P}(x)$. □

For a Bradley–Terry instance $\theta \in \Theta_{B,d}$, let $b = b(\theta)$ denote its coordinate-wise Borda vectors, and write

$$\Delta_i^{\mathrm{B}} := \Delta_i(b(\theta)), \qquad \Delta_i^\theta := \Delta_i(\theta).$$

Recall that

$$\kappa_B = \inf_{|x|\leq 2B} \sigma'(x) = \sigma(2B)\big(1-\sigma(2B)\big) > 0.$$

Lemma B.3 (Borda gaps dominate latent gaps). *If $\theta \in \Theta_{B,d}$, then for every model i,*

$$\Delta_i^{\mathrm{B}} \geq \kappa_B \Delta_i^\theta.$$

Proof. Fix models a, c and an objective r. For every $u \in [K]$, define

$$\bar{b}_{u,r} := \frac{1}{K}\sum_{\ell=1}^{K} \sigma(\theta_{u,r} - \theta_{\ell,r}).$$

For each ℓ, the mean-value theorem gives a point ξ_ℓ between $\theta_{a,r}-\theta_{\ell,r}$ and $\theta_{c,r}-\theta_{\ell,r}$ such that

$$\sigma(\theta_{c,r}-\theta_{\ell,r}) - \sigma(\theta_{a,r}-\theta_{\ell,r}) = \sigma'(\xi_\ell)(\theta_{c,r}-\theta_{a,r}).$$

Hence

$$\bar{b}_{c,r} - \bar{b}_{a,r} = \gamma_{acr}(\theta_{c,r}-\theta_{a,r}), \qquad \gamma_{acr} := \frac{1}{K}\sum_{\ell=1}^{K}\sigma'(\xi_\ell).$$

Because $\theta \in \Theta_{B,d}$, both endpoints defining each ξ_ℓ lie in $[-2B, 2B]$, and therefore $\gamma_{acr} \geq \kappa_B$. Since

$$b_{u,r} = \frac{K\bar{b}_{u,r} - 1/2}{K-1},$$

we obtain

$$b_{c,r} - b_{a,r} = \frac{K}{K-1}\gamma_{acr}(\theta_{c,r}-\theta_{a,r}).$$

In particular, because $K/(K-1) \geq 1$, whenever $\theta_{c,r} \geq \theta_{a,r}$,

$$b_{c,r} - b_{a,r} \geq \kappa_B(\theta_{c,r}-\theta_{a,r}). \tag{9}$$

We then record three consequences.

First, if $m_\theta(a,c) > 0$, then every coordinate difference $\theta_{c,r}-\theta_{a,r}$ is positive, so (9) gives

$$m_b(a,c) \geq \kappa_B m_\theta(a,c). \tag{10}$$

Second, for all a, c,

$$M_b(a,c)^+ \geq \kappa_B M_\theta(a,c)^+. \tag{11}$$

If $M_\theta(a,c) \leq 0$, this is trivial. Otherwise choose an objective attaining $M_\theta(a,c) = \theta_{a,r} - \theta_{c,r} > 0$ and apply (9) with the roles of a and c interchanged.

Third, for every model j,

$$(\Delta_j^\star(b))^+ \geq \kappa_B(\Delta_j^\star(\theta))^+. \tag{12}$$

If $\Delta_j^\star(\theta) \leq 0$, this is trivial. Otherwise choose a model c attaining $\Delta_j^\star(\theta) > 0$ and apply (10) to (j,c).

We now consider the two possible Pareto classes of model i.

If $i \notin \mathcal{P}(\theta)$, coordinate-wise order preservation implies $i \notin \mathcal{P}(b)$. Choose a model c attaining $\Delta_i^\star(\theta) > 0$. Then

$$\Delta_i^{\mathrm{B}} = \Delta_i^\star(b) \geq m_b(i,c) \geq \kappa_B m_\theta(i,c) = \kappa_B \Delta_i^\theta.$$

Suppose instead that $i \in \mathcal{P}(\theta)$. Then $i \in \mathcal{P}(b)$ by Lemma 2.1. Fix $j \neq i$. Since i is Pareto, $M_\theta(i,j) \geq 0$. If $M_\theta(i,j) > 0$, applying (9) at an objective attaining the positive maximum gives

$$M_b(i,j) \geq \kappa_B M_\theta(i,j).$$

If $M_\theta(i,j) = 0$, then some objective satisfies $\theta_{i,r} = \theta_{j,r}$; coordinate-wise order preservation gives $b_{i,r} = b_{j,r}$, and hence again $M_b(i,j) \geq 0 = \kappa_B M_\theta(i,j)$. Thus, in all cases,

$$M_b(i,j) \geq \kappa_B M_\theta(i,j).$$

Moreover, combining (11) and (12) gives

$$M_b(j,i)^+ + (\Delta_j^\star(b))^+ \geq \kappa_B \left[M_\theta(j,i)^+ + (\Delta_j^\star(\theta))^+\right].$$

Therefore, for every $j \neq i$,

$$M_b(i,j) \wedge \left[M_b(j,i)^+ + \left(\Delta_j^\star(b)\right)^+\right] \geq \kappa_B \left\{M_\theta(i,j) \wedge \left[M_\theta(j,i)^+ + \left(\Delta_j^\star(\theta)\right)^+\right]\right\},$$

Taking the minimum over $j \neq i$ yields

$$\Delta_i^{\mathrm{B}} = \delta_i^\star(b) \geq \kappa_B \delta_i^\star(\theta) = \kappa_B \Delta_i^\theta.$$

This proves the claim for every model i. □

B.1.2 High-probability analysis of VB-EGE

We first state the upper-bound argument for a generic vector-Bernoulli mean instance. Fix $b = (b_1, \ldots, b_K) \in ([0,1]^d)^K$. In this section, a query of model i on objective r reveals the next unused observation from an independent stream

$$Z_{i,r,1}, Z_{i,r,2}, \ldots \overset{\text{iid}}{\sim} \mathrm{Bernoulli}(b_{i,r}), \qquad i \in [K],\ r \in [d],$$

These streams are independent of any additional internal randomization used by the algorithm.

Let $\mathbb{P}_b$ denote the induced law of the algorithm and the oracle. For compactness, write

$$\Delta_i^{\mathrm{B}} := \Delta_i(b), \qquad \mathcal{P} := \mathcal{P}(b), \qquad \Delta_{\min}^{\mathrm{B}} := \min_{i \in [K]} \Delta_i^{\mathrm{B}}.$$

Throughout this section, Alg. 1 is run with the common configuration $C = (2,4,4)$ specified in Eq. (7). Let τ count objective-specific binary comparisons, with $\tau = \infty$ if the procedure does not stop. Whenever $\Delta_{\min}^{\mathrm{B}} > 0$, define

$$M_i := 1 + \left\lceil \log_2 \frac{4}{\Delta_i^{\mathrm{B}}} \right\rceil, \qquad L_i(\delta) := \log \frac{4KdM_i^2}{\delta}.$$

Also define the $(1-\delta)$ stopping-time quantile under the vector-Bernoulli instance b by

$$Q_{b,\delta}(\text{VB-EGE}) := \inf\left\{T \in \mathbb{N}_0 : \mathbb{P}_b(\tau \leq T) \geq 1 - \delta\right\}, \qquad \inf \varnothing := \infty.$$

Proposition B.4 (VB-EGE guarantee under vector-Bernoulli feedback). *If $\Delta^{\mathrm{B}}_{\min} > 0$, then Alg. 1, run with the configuration C in (7), satisfies*

$$\mathbb{P}_b\left(\widehat{\mathcal{P}} = \mathcal{P}(b), \quad \tau \le dK + 128d\sum_{i=1}^{K}\frac{L_i(\delta)}{(\Delta_i^{\mathrm{B}})^2}\right) \ge 1-\delta. \tag{13}$$

Consequently,

$$Q_{b,\delta}(\text{VB-EGE}) \le dK + 128d\sum_{i=1}^{K}\frac{L_i(\delta)}{(\Delta_i^{\mathrm{B}})^2} = \widetilde{O}\left(d\sum_{i=1}^{K}(\Delta_i^{\mathrm{B}})^{-2}\right).$$

We first establish several lemmas used in the proof of Prop. B.4. Under (7), the phase schedule is

$$\varepsilon_m = 2^{-m}, \qquad L_m = \log\frac{4Kdm^2}{\delta}, \qquad n_m = \left\lceil\frac{2L_m}{\varepsilon_m^2}\right\rceil, \qquad r_m = \sqrt{\frac{L_m}{2n_m}},$$

and the removal threshold is $4r_m$.

For every phase m, model i, and objective r, define the conceptual phase-m average

$$\widehat{b}_{i,r}^{(m)} := \frac{1}{n_m}\sum_{s=1}^{n_m} Z_{i,r,s}.$$

For an inactive model–objective pair, this quantity is used only in the analysis; the algorithm need not have revealed the corresponding prefix. At an inner-loop iteration with active set A, abbreviate

$$\widehat{m}_A(i,j) := m_{\widehat{b}^{(m)}}(i,j), \quad \widehat{M}_A(i,j) := M_{\widehat{b}^{(m)}}(i,j), \quad \widehat{\Delta}^{\star}_{i,A} := \Delta^{\star}_{i,A}\left(\widehat{b}^{(m)}\right), \quad \widehat{\delta}^{\star}_{i,A} := \delta^{\star}_{i,A}\left(\widehat{b}^{(m)}\right).$$

Lemma B.5 (Empirical gap identities). *For every active set $A \subseteq [K]$ with $|A| \ge 2$ and every $i \in A$,*

$$i \in \widehat{\mathcal{S}}_A \iff \widehat{\Delta}^{\star}_{i,A} \le 0, \qquad i \notin \widehat{\mathcal{S}}_A \iff \widehat{\Delta}^{\star}_{i,A} > 0.$$

and,

$$\widehat{\Delta}_{i,A} = \widehat{\Delta}^{\star}_{i,A} \vee \widehat{\delta}^{\star}_{i,A}.$$

Proof. By definition,

$$\widehat{\Delta}^{\star}_{i,A} = \max_{j\in A\setminus\{i\}} \widehat{m}_A(i,j).$$

Hence $i \notin \widehat{\mathcal{S}}_A$ if and only if some $j \in A\setminus\{i\}$ satisfies $\widehat{m}_A(i,j) > 0$, which is equivalent to $\widehat{\Delta}^{\star}_{i,A} > 0$. The first two equivalences follow.

Suppose first that $i \in \widehat{\mathcal{S}}_A$. Then $\widehat{\Delta}^{\star}_{i,A} \le 0$. Moreover, for every $j \in A\setminus\{i\}$, empirical Pareto membership implies that j does not strictly dominate i. Thus there exists a coordinate $r \in [d]$ such that $\widehat{b}_{i,r} \ge \widehat{b}_{j,r}$, and therefore

$$\widehat{M}_A(i,j) = \max_{r\in[d]}\left(\widehat{b}_{i,r} - \widehat{b}_{j,r}\right) \ge 0.$$

The other term, $\widehat{M}_A(j,i)^+ + \left(\widehat{\Delta}^{\star}_{j,A}\right)^+$, is also nonnegative. Hence every term in the minimum defining $\widehat{\delta}^{\star}_{i,A}$ is nonnegative, so $\widehat{\delta}^{\star}_{i,A} \ge 0 \ge \widehat{\Delta}^{\star}_{i,A}$. Therefore

$$\widehat{\Delta}^{\star}_{i,A} \vee \widehat{\delta}^{\star}_{i,A} = \widehat{\delta}^{\star}_{i,A} = \widehat{\Delta}_{i,A}.$$

Now suppose that $i \notin \widehat{\mathcal{S}}_A$. Then there exists $j \in A\setminus\{i\}$ such that $\widehat{m}_A(i,j) > 0$, or equivalently,

$$\widehat{M}_A(i,j) = -\widehat{m}_A(i,j) < 0.$$

Using this j in the minimum defining $\widehat{\delta}^{\star}_{i,A}$ gives

$$\begin{aligned}\widehat{\delta}^{\star}_{i,A} &\le \widehat{M}_A(i,j) \wedge \left[\widehat{M}_A(j,i)^+ + \left(\widehat{\Delta}^{\star}_{j,A}\right)^+\right] \\ &= \widehat{M}_A(i,j) < 0,\end{aligned}$$

because the second term in the wedge is nonnegative. Since $\widehat{\Delta}^{\star}_{i,A} > 0$, it follows that

$$\widehat{\Delta}^{\star}_{i,A} \vee \widehat{\delta}^{\star}_{i,A} = \widehat{\Delta}^{\star}_{i,A} = \widehat{\Delta}_{i,A}.$$

This proves the claim. □

Lemma B.6 (Uniform Borda concentration)**.** *Define*

$$\mathcal{E} := \left\{ \forall m \geq 1,\ \forall i \in [K],\ \forall r \in [d] : \left| \widehat{b}_{i,r}^{(m)} - b_{i,r} \right| \leq r_m \right\}.$$

Then $\mathbb{P}_b(\mathcal{E}) \geq 1 - \delta$. *Moreover, on* $\mathcal{E}$*, for every phase* m*, active set* A*, and distinct* $i, j \in A$*,*

$$|\widehat{m}_A(i,j) - m_b(i,j)| \leq 2r_m, \qquad \left| \widehat{M}_A(i,j) - M_b(i,j) \right| \leq 2r_m. \tag{14}$$

Proof. For a fixed triple (i, r, m), Hoeffding's inequality and the definition of r_m give

$$\mathbb{P}_b\left(\left| \widehat{b}_{i,r}^{(m)} - b_{i,r} \right| > r_m \right) \leq 2\exp(-2n_m r_m^2) = 2e^{-L_m} = \frac{\delta}{2Kdm^2}.$$

Taking a union bound over $i \in [K]$, $r \in [d]$, and $m \geq 1$ yields

$$\mathbb{P}_b(\mathcal{E}^c) \leq \frac{\delta}{2} \sum_{m=1}^{\infty} \frac{1}{m^2} < \delta.$$

The two margin bounds follow from Lemma B.1. □

For every non-Pareto model $i \notin \mathcal{P}$, fix once and for all a Borda-Pareto witness supplied by Lemma B.2:

$$i^\star \in \arg\max_{p \in \mathcal{P}} m_b(i,p), \qquad m_b(i,i^\star) = \Delta_i^{\mathrm{B}} > 0.$$

For an active set A, define the witness invariant

$$\mathsf{W}(A) := \{ \forall i \notin \mathcal{P}, \quad i \in A \Longrightarrow i^\star \in A \}.$$

Initially, $\mathsf{W}([K])$ holds.

Lemma B.7 (Every removed model is correctly classified)**.** *Assume that* $\mathcal{E}$ *and* $\mathsf{W}(A)$ *hold when a model* $i \in A$ *is removed in phase* m*. Then:*

(i) if $i \notin \widehat{\mathcal{S}}_A$*, then* $i \notin \mathcal{P}$*;*

(ii) if $i \in \widehat{\mathcal{S}}_A$*, then* $i \in \mathcal{P}$*.*

Thus every model removed without acceptance is truly non-Pareto, while every accepted model is truly Pareto.

Proof. For simplicity, we write $r = r_m$.

First suppose $i \notin \widehat{\mathcal{S}}_A$. Since i is removed, $\widehat{\Delta}_{i,A} > 4r$. By Lemma B.5, $\widehat{\Delta}_{i,A} = \widehat{\Delta}^\star_{i,A}$, so there exists $j \in A \setminus \{i\}$ satisfies $\widehat{m}_A(i,j) > 4r$. Using (14),

$$m_b(i,j) \geq \widehat{m}_A(i,j) - 2r > 2r > 0.$$

Hence $i \prec_b j$, and therefore $i \notin \mathcal{P}(b) = \mathcal{P}$.

Now suppose $i \in \widehat{\mathcal{S}}_A$ but, toward a contradiction, $i \notin \mathcal{P}$. The witness invariant gives $i^\star \in A$, and

$$m_b(i,i^\star) = \Delta_i^{\mathrm{B}} > 0, \qquad M_b(i,i^\star) = -\Delta_i^{\mathrm{B}}.$$

If $\Delta_i^{\mathrm{B}} > 2r$, then (14) implies

$$\widehat{m}_A(i,i^\star) \geq \Delta_i^{\mathrm{B}} - 2r > 0,$$

contradicting $i \in \widehat{\mathcal{S}}_A$. Thus necessarily $\Delta_i^{\mathrm{B}} \leq 2r$.

Since $i \in \widehat{\mathcal{S}}_A$, its empirical classification gap is $\widehat{\Delta}_{i,A} = \widehat{\delta}^\star_{i,A}$. Taking $j = i^\star$ in the defining minimum gives

$$\begin{aligned}
\widehat{\Delta}_{i,A} &\leq \widehat{M}_A(i,i^\star) \wedge \left[\widehat{M}_A(i^\star,i)^+ + \left(\widehat{\Delta}^\star_{i^\star,A}\right)^+ \right] \\
&\leq \widehat{M}_A(i,i^\star) \\
&\leq M_b(i,i^\star) + 2r \\
&= -\Delta_i^{\mathrm{B}} + 2r \leq 2r,
\end{aligned}$$

contradicting the removal condition $\widehat{\Delta}_{i,A} > 4r$. Therefore $i \in \mathcal{P}$. □

Lemma B.8 (Witness-invariant preservation)**.** *Assume that $\mathcal{E}$ and $\mathsf{W}(A)$ hold before a removal in phase m. After the removal, the new active set A' also satisfies $\mathsf{W}(A')$.*

Proof. Write $r = r_m$. By Lemma B.7, the only possible way to violate the invariant is to remove a true Pareto model $p \in \mathcal{P}$ while leaving active a non-Pareto model $i \notin \mathcal{P}$ whose fixed witness is $i^\star = p$. Suppose, toward a contradiction, that this occurs. Immediately before the removal, $p, i \in A$. Since the removed model p is truly Pareto, Lemma B.7 implies that it cannot be empirically non-Pareto. Thus $p \in \widehat{\mathcal{S}}_A$, and

$$\widehat{\Delta}_{p,A} = \widehat{\delta}^\star_{p,A} > 4r. \tag{15}$$

There are two cases to be considered.

Case 1: p empirically strictly dominates i. Then $\widehat{m}_A(i,p) > 0$, so $i \notin \widehat{\mathcal{S}}_A$ and $\widehat{\Delta}_{i,A} = \widehat{\Delta}^\star_{i,A} > 0$. Moreover, $\widehat{M}_A(i,p)^+ = 0$. Taking $j = i$ in the minimum defining $\widehat{\delta}^\star_{p,A}$ yields

$$\widehat{\Delta}_{p,A} \le \widehat{M}_A(i,p)^+ + (\widehat{\Delta}^\star_{i,A})^+ = \widehat{\Delta}_{i,A}.$$

Since p is the model selected for removal, it belongs to $G := \arg\max_{u \in A} \widehat{\Delta}_{u,A}$. If the preceding inequality is strict, then $\widehat{\Delta}_{i,A} > \widehat{\Delta}_{p,A}$, contradicting $p \in G$. If equality holds, then $i \in G \setminus \widehat{\mathcal{S}}_A$, so the non-Pareto-first tie-breaking rule Rem. 2.1 selects an empirical non-Pareto maximizer rather than p. In either case, the algorithm cannot remove p, which results in a contradiction.

Case 2: p does not empirically strictly dominate i. Then $\widehat{m}_A(i,p) \le 0$, and hence

$$\widehat{M}_A(i,p) = -\widehat{m}_A(i,p) \ge 0.$$

Since $M_b(i,p) = -m_b(i,p) = -\Delta^{\mathrm{B}}_i$, the uniform margin deviation (14) gives

$$0 \le \widehat{M}_A(i,p) = \widehat{M}_A(i,p)^+ \le 2r - \Delta^{\mathrm{B}}_i.$$

So $\Delta^{\mathrm{B}}_i \le 2r$. Moreover, for every $q \in A \setminus \{i\}$,

$$\widehat{m}_A(i,q) \le m_b(i,q) + 2r \le \Delta^{\mathrm{B}}_i + 2r,$$

Therefore,

$$\left(\widehat{\Delta}^\star_{i,A}\right)^+ \le \Delta^{\mathrm{B}}_i + 2r.$$

Since $p \in \widehat{\mathcal{S}}_A$, $\widehat{\Delta}_{p,A} = \widehat{\delta}^\star_{p,A}$. Taking $j = i$ in the defining minimum gives

$$\begin{aligned}
\widehat{\Delta}_{p,A} &\le \widehat{M}_A(p,i) \wedge \left[\widehat{M}_A(i,p)^+ + \left(\widehat{\Delta}^\star_{i,A}\right)^+\right] \\
&\le \widehat{M}_A(i,p)^+ + \left(\widehat{\Delta}^\star_{i,A}\right)^+ \\
&\le \left(2r - \Delta^{\mathrm{B}}_i\right) + \left(\Delta^{\mathrm{B}}_i + 2r\right) \\
&= 4r,
\end{aligned}$$

contradicting (15). Hence the invariant is preserved. □

Proposition B.9 (Correctness on the confidence event)**.** *On $\mathcal{E}$, Alg. 1 never misclassifies a removed model. If it stops, its recommendation equals $\mathcal{P}(b)$.*

Proof. The invariant holds initially and is preserved after every removal by Lemma B.8. By induction over removals and Lemma B.7, every model added to B belongs to $\mathcal{P}$, while every model removed without being added to B lies outside $\mathcal{P}$.

If the algorithm stops with a singleton active set $A = \{a\}$ and $a \notin \mathcal{P}$, then the invariant would require the distinct witness $a^\star$ to remain active, contradicting $|A| = 1$. Thus the final active model is Pareto. Every model outside $B \cup A$ was previously removed without acceptance and hence lies outside $\mathcal{P}$. Conversely, every model in $B \cup A$ is Pareto. Therefore the returned set $B \cup A$ equals $\mathcal{P}(b)$. □

Lemma B.10 (Removal once the radius is small)**.** *Assume $\mathcal{E}$. Consider a phase m and a model i that is active at the beginning of the inner loop. If*

$$r_m < \frac{\Delta_i^{\mathrm{B}}}{8},$$

then either the algorithm stops during phase m, or model i is removed during that phase.

Proof. Write $r = r_m$, and suppose that i remains active at some inner-loop iteration with $|A| \geq 2$. By Lemma B.8, the witness invariant $\mathsf{W}(A)$ holds.

If $i \notin \mathcal{P}$, then its fixed witness $i^\star$ remains active. By (14),

$$\widehat{m}_A(i, i^\star) \geq \Delta_i^{\mathrm{B}} - 2r > 0.$$

Thus $i \notin \widehat{\mathcal{S}}_A$, and Lemma B.5 gives

$$\widehat{\Delta}_{i,A} = \widehat{\Delta}^\star_{i,A} \geq \Delta_i^{\mathrm{B}} - 2r > 6r > 4r.$$

Now suppose $i \in \mathcal{P}$. Then $\Delta_i^{\mathrm{B}} = \delta_i^\star(b)$. Hence, for every $j \in A \setminus \{i\}$, both components in the defining wedge are at least Δ_i^{B}:

$$M_b(i,j) \geq \Delta_i^{\mathrm{B}}, \qquad M_b(j,i)^+ + (\Delta_j^\star(b))^+ \geq \Delta_i^{\mathrm{B}}. \tag{16}$$

By (14),

$$\widehat{M}_A(i,j) \geq \Delta_i^{\mathrm{B}} - 2r > 0.$$

Thus no active model empirically strictly dominates i, so $i \in \widehat{\mathcal{S}}_A$. For any active model $j \in A$,

$$\big(\widehat{\Delta}^\star_{j,A}\big)^+ \geq \big(\Delta_j^\star(b)\big)^+ - 2r.$$

Indeed, if $j \in \mathcal{P}$, then $\Delta_j^\star(b) \leq 0$, so the claim is immediate since the left-hand side is nonnegative. If $j \notin \mathcal{P}$, the witness invariant ensures that its fixed Pareto witness $j^\star$ remains in A. Hence

$$\widehat{\Delta}^\star_{j,A} = \max_{q \in A \setminus \{j\}} \widehat{m}_A(j,q) \geq \widehat{m}_A(j, j^\star) \geq m_b(j, j^\star) - 2r = \Delta_j^\star(b) - 2r.$$

Since $\Delta_j^\star(b) > 0$ in this case,

$$\big(\widehat{\Delta}^\star_{j,A}\big)^+ \geq \big(\Delta_j^\star(b)\big)^+ - 2r.$$

Combining with (16) yields

$$\widehat{M}_A(j,i)^+ + (\widehat{\Delta}^\star_{j,A})^+ \geq \Delta_i^{\mathrm{B}} - 4r > 4r.$$

Together with $\widehat{M}_A(i,j) \geq \Delta_i^{\mathrm{B}} - 2r$, this shows that every bracket in the definition of $\widehat{\delta}^\star_{i,A}$ is at least $\Delta_i^{\mathrm{B}} - 4r$. Therefore

$$\widehat{\Delta}_{i,A} = \widehat{\delta}^\star_{i,A} \geq \Delta_i^{\mathrm{B}} - 4r > 4r.$$

In either case, as long as i remains active and $|A| \geq 2$, its empirical gap exceeds the removal threshold. Hence the inner loop cannot terminate while retaining i: the algorithm must remove i, unless it stops first because only one active model remains. □

Proof of Prop. B.4. By Lemma B.6, $\mathbb{P}_b(\mathcal{E}) \geq 1 - \delta$. Work on $\mathcal{E}$. Correctness upon stopping follows from Prop. B.9; it remains to bound the query count.

For each model i, define

$$m_i := \min\left\{ m \geq 1 : \varepsilon_m < \frac{\Delta_i^{\mathrm{B}}}{4} \right\}.$$

This integer is well defined because $\varepsilon_m \downarrow 0$. Moreover, $0 < \Delta_i^{\mathrm{B}} \leq 1$: for a non-Pareto model the raw domination gap is at most 1, while for a Pareto model the defining wedge is bounded above by $M_b(i,j) \leq 1$. Hence $m_i > 1$. By minimality of m_i,

$$\varepsilon_{m_i - 1} \geq \frac{\Delta_i^{\mathrm{B}}}{4}, \qquad \varepsilon_{m_i} = \frac{1}{2}\varepsilon_{m_i - 1} \geq \frac{\Delta_i^{\mathrm{B}}}{8}.$$

On the other hand, the defining property of m_i gives

$$\varepsilon_{m_i} < \frac{\Delta_i^{\mathrm{B}}}{4}.$$

Since $n_m \geq 2L_m/\varepsilon_m^2$,

$$r_m = \sqrt{\frac{L_m}{2n_m}} \leq \frac{\varepsilon_m}{2}.$$

Therefore

$$r_{m_i} < \frac{\Delta_i^{\mathrm{B}}}{8}.$$

By Lemma B.10, if the algorithm reaches phase m_i and model i is still active, then during that phase either the algorithm stops or i is removed. Thus no coordinate of model i is sampled after phase m_i.

In particular, the algorithm stops after finitely many phases. Indeed, let

$$\overline{m} := \max_{i \in [K]} m_i < \infty.$$

If the algorithm had not stopped by the end of phase $\overline{m}$, any model still active at that time would have remained active beyond its phase m_i, contradicting Lemma B.10.

Consequently, the algorithm stops by phase $\overline{m}$, and model i is never sampled after phase m_i. Each of its d coordinates therefore receives at most n_{m_i} observations, so

$$\tau \leq d \sum_{i=1}^{K} n_{m_i}.$$

It remains to bound n_{m_i}. By its definition and the lower bound $\varepsilon_{m_i} \geq \Delta_i^{\mathrm{B}}/8$,

$$n_{m_i} \leq 1 + \frac{2}{\varepsilon_{m_i}^2} \log \frac{4Kdm_i^2}{\delta} \leq 1 + \frac{128}{(\Delta_i^{\mathrm{B}})^2} \log \frac{4Kdm_i^2}{\delta}.$$

Furthermore, $m_i \leq M_i$, because

$$\varepsilon_{M_i} = 2^{-M_i} \leq \frac{\Delta_i^{\mathrm{B}}}{8} < \frac{\Delta_i^{\mathrm{B}}}{4}.$$

Consequently,

$$n_{m_i} \leq 1 + \frac{128 L_i(\delta)}{(\Delta_i^{\mathrm{B}})^2}.$$

Combining the preceding two displays gives

$$\tau \leq dK + 128d \sum_{i=1}^{K} \frac{L_i(\delta)}{(\Delta_i^{\mathrm{B}})^2}.$$

Thus, on $\mathcal{E}$, the algorithm returns $\mathcal{P}(b)$ within the stated query budget. Since $\mathbb{P}_b(\mathcal{E}) \geq 1 - \delta$, (13) follows. The quantile statement follows directly from the definition of $Q_{b,\delta}(\text{VB-EGE})$. □

B.1.3 Change-of-measure tools

Lemma B.11 (Monotonicity of binary relative entropy)**.** *Let*

$$0 \leq q \leq b < a \leq p \leq 1.$$

Then

$$\mathrm{kl}(p, q) \geq \mathrm{kl}(a, b),$$

where binary relative entropy is extended to the boundary by the usual conventions.

Proof. For $p, q \in (0,1)$,

$$\frac{\partial}{\partial p}\,\mathrm{kl}(p,q) = \log\frac{p(1-q)}{q(1-p)},$$

which is positive whenever $p > q$. Moreover,

$$\frac{\partial}{\partial q}\,\mathrm{kl}(p,q) = \frac{q-p}{q(1-q)},$$

which is negative whenever $p > q$. Thus $\mathrm{kl}(p,q)$ is increasing in its first argument and decreasing in its second argument on the region $p > q$. The boundary cases follow by taking the corresponding one-sided limits; cases in which the left-hand side is infinite are immediate. □

Lemma B.12 (Logistic KL upper bound)**.** *For all $x, y \in \mathbb{R}$,*

$$\mathrm{kl}(\sigma(x), \sigma(y)) \le \frac{(x-y)^2}{8}.$$

Proof. The Bernoulli distribution with mean $\sigma(u)$ is the one-parameter exponential family with log-partition function $A(u) = \log(1+e^u)$. Since $A''(u) = \sigma(u)(1-\sigma(u)) \le 1/4$, the function A is globally $1/4$-smooth. The corresponding Bregman-divergence representation gives

$$\mathrm{kl}(\sigma(x), \sigma(y)) = A(y) - A(x) - A'(x)(y-x) \le \frac{(y-x)^2}{8}.$$

□

For $1 \le i < j \le K$, objective $r \in [d]$, and $u \in \mathbb{N}_0$ with $u \le \tau$, define

$$N^{\mathrm{pair}}_{ij,r}(u) := \sum_{t=1}^{u} \mathbf{1}\{R_t = r,\ \{I_t, J_t\} = \{i,j\}\}.$$

Thus $N^{\mathrm{pair}}_{ij,r}(u)$ counts comparisons of the unordered pair $\{i,j\}$ on objective r among the first u queries.

For every model $i \in [K]$, define its incident-comparison count by

$$N^{\mathrm{inc}}_{i,r}(u) := \sum_{j<i} N^{\mathrm{pair}}_{ji,r}(u) + \sum_{j>i} N^{\mathrm{pair}}_{ij,r}(u).$$

Lemma B.13 (Bounded-horizon transportation)**.** *Fix $T \in \mathbb{N}_0$. Let θ and λ be two instances, and let G be an event measurable with respect to the transcript stopped at $\tau \wedge T$, including the terminal recommendation if the algorithm stops by time T. Then*

$$\sum_{r=1}^{d} \sum_{1 \le i < j \le K} \mathbb{E}_\theta\left[N^{\mathrm{pair}}_{ij,r}(\tau \wedge T)\right] \mathrm{kl}\Big(p^{(r)}_{ij}(\theta), p^{(r)}_{ij}(\lambda)\Big) \ge \mathrm{kl}(\mathbb{P}_\theta(G), \mathbb{P}_\lambda(G)). \tag{17}$$

Proof. Let $\mathbb{P}^T_\theta$ and $\mathbb{P}^T_\lambda$ denote the laws of the transcript stopped at $\tau \wedge T$. View the stopped transcript as a length-T adaptive experiment with a common cemetery action and observation after stopping, which contributes zero KL divergence. At each time, the query is predictable from the past and the algorithmic randomization, whose law is the same under both instances. The chain rule for KL divergence therefore gives

$$\mathrm{KL}(\mathbb{P}^T_\theta, \mathbb{P}^T_\lambda) = \mathbb{E}_\theta\left[\sum_{t=1}^{\tau\wedge T} \mathrm{kl}\left(p^{(R_t)}_{I_t J_t}(\theta), p^{(R_t)}_{I_t J_t}(\lambda)\right)\right].$$

For a reverse-orientation query, $p^{(r)}_{ji} = 1 - p^{(r)}_{ij}$, and $\mathrm{kl}(1-p, 1-q) = \mathrm{kl}(p,q)$. Grouping both orientations by unordered pair turns the right-hand side into the left-hand side of (17). Finally, applying data processing to the binary statistic $\mathbf{1}_G$ gives

$$\mathrm{KL}\big(\mathbb{P}^T_\theta \,\big\|\, \mathbb{P}^T_\lambda\big) \ge \mathrm{kl}\left(\mathbb{P}_\theta(G), \mathbb{P}_\lambda(G)\right),$$

which proves the claim. □

B.2 Proofs of main results

B.2.1 Proof of Lemma 2.1

Proof of Lemma 2.1. For every model $u \in [K]$ and objective $r \in [d]$, define the self-included Borda coordinate

$$\bar{b}_{u,r} := \frac{1}{K} \sum_{\ell=1}^{K} \sigma(\theta_{u,r} - \theta_{\ell,r}).$$

Since $\sigma(0) = 1/2$,

$$b_{u,r} = \frac{K\bar{b}_{u,r} - 1/2}{K-1}.$$

In particular, $b_{u,r}$ is a strictly increasing affine function of $\bar{b}_{u,r}$.

Fix $i, j \in [K]$ and $r \in [d]$. Suppose first that $\theta_{i,r} > \theta_{j,r}$. Then, for every $\ell \in [K]$,

$$\theta_{i,r} - \theta_{\ell,r} > \theta_{j,r} - \theta_{\ell,r}.$$

Because the logistic function is strictly increasing,

$$\sigma(\theta_{i,r} - \theta_{\ell,r}) > \sigma(\theta_{j,r} - \theta_{\ell,r}) \qquad \text{for every } \ell \in [K].$$

Averaging over ℓ yields $\bar{b}_{i,r} > \bar{b}_{j,r}$, and hence $b_{i,r} > b_{j,r}$.

Exchanging i and j proves that $\theta_{i,r} < \theta_{j,r}$ implies $b_{i,r} < b_{j,r}$. If $\theta_{i,r} = \theta_{j,r}$, then the summands defining $\bar{b}_{i,r}$ and $\bar{b}_{j,r}$ coincide for every ℓ, so $b_{i,r} = b_{j,r}$. Therefore,

$$\operatorname{sign}(b_{i,r} - b_{j,r}) = \operatorname{sign}(\theta_{i,r} - \theta_{j,r}),$$

which proves (2).

Applying the sign identity coordinate-wise gives

$$\theta_{i,r} < \theta_{j,r} \text{ for all } r \in [d] \quad \Longleftrightarrow \quad b_{i,r} < b_{j,r} \text{ for all } r \in [d].$$

Hence $i \prec_\theta j$ if and only if $i \prec_b j$. Since the strict Pareto frontier is determined entirely by this domination relation,

$$\mathcal{P}(b) = \mathcal{P}(\theta) = \mathcal{P}^\star,$$

which proves (3). □

B.2.2 Proof of Thm. 3.1

Proof of Thm. 3.1. Conditional on $\mathcal{F}_{t-1}$ and $(I_t, R_t) = (i, r)$, the opponent J_t is uniform on $[K]\backslash\{i\}$. By (1),

$$\mathbb{P}_\theta(Y_t = 1 \mid \mathcal{F}_{t-1}, I_t = i, R_t = r) = \frac{1}{K-1} \sum_{j \neq i} p_{ij}^{(r)}(\theta) = b_{i,r}(\theta).$$

Since the opponent and comparison randomness are fresh at every query, the oracle can be realized by mutually independent pre-generated streams

$$Z_{i,r,1}, Z_{i,r,2}, \ldots \overset{\text{iid}}{\sim} \text{Bernoulli}(b_{i,r}(\theta)),$$

revealing the next unused observation whenever the algorithm queries (i, r). Alg. 1 discards the opponent identities and uses only these binary outcomes, so its stopping time and recommendation have the same law under this stream representation. Hence Prop. B.4 applies with $b = b(\theta)$ under $\mathbb{P}_\theta$. By Lemma 2.1,

$$\mathcal{P}(b(\theta)) = \mathcal{P}(\theta).$$

Let M_i and $L_i(\delta)$ be the Borda-space quantities from App. B.1.2. By Lemma B.3, $\Delta_i^{\mathrm{B}} \geq \kappa_B \Delta_i^\theta$, and therefore

$$M_i = 1 + \left\lceil \log_2 \frac{4}{\Delta_i^{\mathrm{B}}} \right\rceil \leq 1 + \left\lceil \log_2 \frac{4}{\kappa_B \Delta_i^\theta} \right\rceil = M_i^\theta.$$

It follows that $L_i(\delta) \le L_i^\theta(\delta)$ and

$$\frac{L_i(\delta)}{(\Delta_i^{\mathrm{B}})^2} \le \frac{L_i^\theta(\delta)}{\kappa_B^2(\Delta_i^\theta)^2}.$$

Substituting this inequality into (13), with $b = b(\theta)$, gives

$$\mathbb{P}_\theta\left(\widehat{\mathcal{P}} = \mathcal{P}(\theta), \quad \tau \le dK + \frac{128d}{\kappa_B^2}\sum_{i=1}^{K}\frac{L_i^\theta(\delta)}{(\Delta_i^\theta)^2}\right) \ge 1-\delta.$$

The displayed event gives the stated bound on $Q_{\theta,\delta}(\text{VB-EGE})$. Absorbing the logarithmic factors $L_i^\theta(\delta)$ yields the $\widetilde{O}$ expression. If B is bounded by an absolute constant, i.e., $B = O(1)$, then κ_B is bounded below by a positive absolute constant, i.e., $\kappa_B = \Theta(1)$. □

B.2.3 Proof of Thm. 3.2

We first establish a quantile lower bound for a symmetric instance and then apply it to the gap-separated class.

Fix $\Delta > 0$ and consider the symmetric instance θ^Δ defined by

$$\theta^\Delta_{1,r} = 0, \qquad \theta^\Delta_{i,r} = -\Delta, \qquad i \in \{2,\dots,K\},\ r \in [d].$$

The first model strictly dominates every other model, and hence $\mathcal{P}(\theta^\Delta) = \{1\}$.

To obtain a lower bound, we construct alternatives under which one of the dominated models becomes Pareto. For $i \in \{2,\dots,K\}$, $r \in [d]$, and $\eta > 0$, define the one-coordinate alternative

$$\lambda^{i,r,\eta}_{a,s} := \begin{cases} \eta, & (a,s) = (i,r), \\ \theta^\Delta_{a,s}, & \text{otherwise.} \end{cases}$$

The alternative changes only the ith model's score on objective r, moving it from $-\Delta$ to $\eta > 0$. Consequently, model i is Pareto under $\lambda^{i,r,\eta}$, so that $\mathcal{P}(\lambda^{i,r,\eta}) \ne \mathcal{P}(\theta^\Delta)$.

The parameter η measures how far the perturbed coordinate lies beyond the Pareto decision boundary at 0. Smaller η yields a harder alternative, since the Pareto frontier still changes while the largest affected Bradley–Terry logit shift, $\Delta + \eta$, becomes smaller.

Lemma B.14 (Quantile lower bound)**.** *Let $\delta \in (0, 1/2)$, and let A be δ-correct over an instance class $\mathcal{C}$ containing θ^Δ. Suppose that there exists $\bar\eta > 0$ such that, for every $i \in \{2,\dots,K\}$ and $r \in [d]$, the class $\mathcal{C}$ contains at least one alternative $\lambda^{i,r,\eta_{i,r}}$ with $0 < \eta_{i,r} \le \bar\eta$. Then, for every $\alpha \in (0, 1-2\delta)$,*

$$Q_{\theta^\Delta,\alpha}(\mathsf{A}) \ge \frac{4d(K-1)}{(\Delta+\bar\eta)^2}\,\mathrm{kl}(1-\delta-\alpha,\delta).$$

If, moreover, arbitrarily small frontier-changing perturbations are admissible, namely, for every $i \in \{2,\dots,K\}$ and $r \in [d]$ there exists a sequence $\eta_n^{i,r} \downarrow 0$ such that $\lambda^{i,r,\eta_n^{i,r}} \in \mathcal{C}$ for every n, then

$$Q_{\theta^\Delta,\alpha}(\mathsf{A}) \ge \frac{4d(K-1)}{\Delta^2}\,\mathrm{kl}(1-\delta-\alpha,\delta).$$

In particular, if $\alpha = \delta < 1/3$, then

$$Q_{\theta^\Delta,\delta}(\mathsf{A}) \ge \frac{4d(K-1)}{\Delta^2}\,\mathrm{kl}(1-2\delta,\delta) = \Omega\left(\frac{dK}{\Delta^2}\log\frac{1}{\delta}\right) \qquad \text{as } \delta \downarrow 0.$$

The lower bound holds under the full adaptive ordered-pair oracle of Sec. 1.1, and therefore also applies to the uniform-opponent implementation used by VB-EGE. The proof uses a bounded-horizon change-of-measure argument (Kaufmann et al., 2016, Lemma 1).

Proof of Lemma B.14. It suffices to lower-bound every finite T satisfying

$$\mathbb{P}_{\theta^\Delta}(\tau \le T) \ge 1-\alpha. \tag{18}$$

If no such T exists, the quantile is infinite and the claim is immediate.

Fix a suboptimal model $i \in \{2, \dots, K\}$, objective $r \in [d]$, and an admissible $\eta > 0$. Under θ^{Δ}, model 1 strictly dominates all others, so $\mathcal{P}(\theta^{\Delta}) = \{1\}$. Under $\lambda^{i,r,\eta}$, model 1 does not dominate i, because $\lambda^{i,r,\eta}_{i,r} = \eta > 0 = \lambda^{i,r,\eta}_{1,r}$. No other model dominates i, since its coordinate r is smaller than η. Thus i is Pareto under the alternative and

$$\mathcal{P}(\lambda^{i,r,\eta}) \neq \{1\} = \mathcal{P}(\theta^{\Delta}). \tag{19}$$

Consider the stopped event

$$G_T := \{\tau \leq T,\ \widehat{\mathcal{P}} = \{1\}\}.$$

By δ-correctness and (18),

$$\mathbb{P}_{\theta^{\Delta}}(G_T) \geq 1 - \delta - \alpha.$$

By (19), outputting $\{1\}$ under $\lambda^{i,r,\eta}$ is an error, so

$$\mathbb{P}_{\lambda^{i,r,\eta}}(G_T) \leq \delta.$$

Since

$$\mathbb{P}_{\theta^{\Delta}}(G_T) \geq 1 - \delta - \alpha > \delta \geq \mathbb{P}_{\lambda^{i,r,\eta}}(G_T),$$

Lemma B.11 yields

$$\mathrm{kl}(\mathbb{P}_{\theta^{\Delta}}(G_T), \mathbb{P}_{\lambda^{i,r,\eta}}(G_T)) \geq \ell^{\mathrm{hp}}_{\delta,\alpha} := \mathrm{kl}(1 - \delta - \alpha, \delta).$$

The two instances differ only in objective r of model i. Thus all queries outside objective r, or not involving i, contribute zero KL in Lemma B.13. For a comparison in objective r involving i, the Bradley–Terry logit changes by at most $\Delta + \eta$: it changes from $-\Delta$ to η against model 1, and from 0 to $\Delta + \eta$ against any other suboptimal model. By Lemma B.12, every affected comparison contributes at most $(\Delta + \eta)^2/8$. Combining Lemma B.13 with Sec. B.2.3 yields

$$\ell^{\mathrm{hp}}_{\delta,\alpha} \leq \frac{(\Delta + \eta)^2}{8} \mathbb{E}_{\theta^{\Delta}}\left[N^{\mathrm{inc}}_{i,r}(\tau \wedge T)\right].$$

Therefore

$$\mathbb{E}_{\theta^{\Delta}}\left[N^{\mathrm{inc}}_{i,r}(\tau \wedge T)\right] \geq \frac{8\ell^{\mathrm{hp}}_{\delta,\alpha}}{(\Delta + \eta)^2}.$$

By the assumption of Lemma B.14, for every $i \in \{2, \dots, K\}$ and $r \in [d]$, one can choose $\eta_{i,r} \in (0, \bar{\eta}]$ such that $\lambda^{i,r,\eta_{i,r}} \in \mathcal{C}$. Therefore,

$$\mathbb{E}_{\theta^{\Delta}}\left[N^{\mathrm{inc}}_{i,r}(\tau \wedge T)\right] \geq \frac{8\ell^{\mathrm{hp}}_{\delta,\alpha}}{(\Delta + \eta_{i,r})^2} \geq \frac{8\ell^{\mathrm{hp}}_{\delta,\alpha}}{(\Delta + \bar{\eta})^2}.$$

Summing over $i \in \{2, \dots, K\}$ and $r \in [d]$ gives

$$\sum_{i=2}^{K} \sum_{r=1}^{d} \mathbb{E}_{\theta^{\Delta}}\left[N^{\mathrm{inc}}_{i,r}(\tau \wedge T)\right] \geq \frac{8d(K-1)\ell^{\mathrm{hp}}_{\delta,\alpha}}{(\Delta + \bar{\eta})^2}.$$

On every sample path,

$$\sum_{i=2}^{K} \sum_{r=1}^{d} N^{\mathrm{inc}}_{i,r}(\tau \wedge T) \leq 2(\tau \wedge T) \leq 2T,$$

because one objective-specific binary comparison is incident to at most two models in $\{2, \dots, K\}$. Combining the two bounds gives

$$T \geq \frac{4d(K-1)}{(\Delta + \bar{\eta})^2} \mathrm{kl}(1 - \delta - \alpha, \delta).$$

Taking the infimum over all finite T satisfying

$$\mathbb{P}_{\theta^{\Delta}}(\tau \leq T) \geq 1 - \alpha$$

proves the first assertion.

The sequence assumption in the second part of Lemma B.14 implies the preceding admissibility condition for every $\bar{\eta} > 0$. Applying the first bound with arbitrary $\bar{\eta} > 0$ and letting $\bar{\eta} \downarrow 0$ gives

$$Q_{\theta^{\Delta},\alpha}(\mathsf{A}) \geq \frac{4d(K-1)}{\Delta^2} \operatorname{kl}(1-\delta-\alpha, \delta).$$

The specialization $\alpha = \delta < 1/3$ follows from

$$\operatorname{kl}(1-2\delta, \delta) = \Theta\left(\log \frac{1}{\delta}\right) \qquad \text{as } \delta \downarrow 0.$$

□

Remark B.1 (Admissible alternatives). For the bounded-range class $\Theta_{B,d}$, the small-perturbation condition in the second part of Lemma B.14 holds whenever $0 < \Delta < 2B$. For example, one may take $\eta_n := \frac{2B-\Delta}{n+1} \downarrow 0$.

This condition cannot hold on the gap-separated subclass $\Theta_{B,d}(\Delta)$. Indeed, for every $\eta > 0$, $\Delta_i\big(\lambda^{i,r,\eta}\big) \leq \eta$, so the alternatives eventually leave $\Theta_{B,d}(\Delta)$ as $\eta \downarrow 0$. However, if $\Delta \leq B$, then

$$\theta^{\Delta}, \lambda^{i,r,\Delta} \in \Theta_{B,d}(\Delta) \qquad \text{for every } i \geq 2,\ r \in [d].$$

The first part of Lemma B.14 therefore applies with $\bar{\eta} = \Delta$, which is the construction used in Thm. 3.2.

We now apply the preceding lemma to prove the minimax lower bound.

Proof of Thm. 3.2. Let A be any algorithm that is δ-correct on $\Theta_{B,d}(\Delta)$, and consider the symmetric instance θ^{Δ} from Lemma B.14. Directly from Def. 2.2,

$$\mathcal{P}(\theta^{\Delta}) = \{1\}, \qquad \Delta_a(\theta^{\Delta}) = \Delta \quad \text{for every } a \in [K].$$

For every $i \geq 2$ and $r \in [d]$, let $\lambda^{i,r,\Delta}$ replace only $\theta^{\Delta}_{i,r} = -\Delta$ by $+\Delta$. A direct calculation gives

$$\mathcal{P}(\lambda^{i,r,\Delta}) = \begin{cases} \{i\}, & d = 1, \\ \{1, i\}, & d \geq 2, \end{cases}$$

and

$$\Delta_1(\lambda^{i,r,\Delta}) = \Delta_i(\lambda^{i,r,\Delta}) = \Delta, \qquad \Delta_k(\lambda^{i,r,\Delta}) \geq \Delta \quad \text{for every } k \notin \{1, i\}.$$

Moreover,

$$\max_{s\in[d]} \max_{a,b\in[K]} |\theta^{\Delta}_{a,s} - \theta^{\Delta}_{b,s}| = \Delta,$$

whereas

$$\max_{s\in[d]} \max_{a,b\in[K]} |\lambda^{i,r,\Delta}_{a,s} - \lambda^{i,r,\Delta}_{b,s}| \leq 2\Delta.$$

Since $\Delta \leq B$,

$$\theta^{\Delta}, \lambda^{i,r,\Delta} \in \Theta_{B,d}(\Delta).$$

Since $\lambda^{i,r,\Delta} \in \Theta_{B,d}(\Delta)$ for every $i \geq 2$ and $r \in [d]$, the first part of Lemma B.14 applies with $\bar{\eta} = \Delta$ and $\alpha = \delta$. Hence

$$\sup_{\theta\in\Theta_{B,d}(\Delta)} Q_{\theta,\delta}(\mathsf{A}) \geq Q_{\theta^{\Delta},\delta}(\mathsf{A}) \geq \frac{d(K-1)}{\Delta^2} \operatorname{kl}(1-2\delta, \delta).$$

Taking the infimum over all such A proves the minimax lower bound. □

B.2.4 Proof of Cor. 3.3

Proof of Cor. 3.3. For every $\theta \in \Theta_{B,d}(\Delta)$,

$$\Delta_i^\theta \geq \Delta, \qquad M_i^\theta \leq \left\lceil \log_2 \frac{8}{\kappa_B \Delta} \right\rceil,$$
$$L_i^\theta(\delta) \leq \log \frac{4Kd\left(\left\lceil \log_2 \frac{8}{\kappa_B \Delta} \right\rceil\right)^2}{\delta}.$$

Hence Thm. 3.1 gives

$$\sup_{\theta \in \Theta_{B,d}(\Delta)} Q_{\theta,\delta}(\text{VB-EGE}) \leq dK + \frac{128dK}{\kappa_B^2 \Delta^2} \log \frac{4Kd\left(\left\lceil \log_2 \frac{8}{\kappa_B \Delta} \right\rceil\right)^2}{\delta}.$$

The joint recovery statement in Thm. 3.1 also shows that VB-EGE is δ-correct on $\Theta_{B,d}(\Delta)$. It is therefore admissible in the infimum defining $\mathbf{Q}_{B,d}(\Delta, \delta)$, and hence

$$\mathbf{Q}_{B,d}(\Delta, \delta) \leq \sup_{\theta \in \Theta_{B,d}(\Delta)} Q_{\theta,\delta}(\text{VB-EGE}).$$

Combining this upper bound with Thm. 3.2 gives the minimax rate. Since $K \geq 2$, we have $K - 1 \geq K/2$. Finally, for fixed B, κ_B is bounded below by a positive constant, while

$$\mathrm{kl}(1 - 2\delta, \delta) = \Theta\left(\log \frac{1}{\delta}\right) \qquad \text{as } \delta \downarrow 0.$$

Moreover, because $\Delta \leq 1$ and $\delta < 1/3$, the additive dK term is absorbed by the displayed upper-bound term. Hence VB-EGE matches the polynomial dependence $dK\Delta^{-2}$ up to logarithmic factors in K, d, and Δ^{-1}, and matches the $\log(1/\delta)$ confidence dependence. □

C Additional details and results for synthetic experiments

C.1 Details for the data-generating process

In the scaling experiments, the parameter grids are specified in Tab. 1.

Table 1: Parameters used in the scaling experiments.

Study	Parameter grid
Scaling in K	$K \in \{16, 32, 64, 128\}$ at $d = 4, \Delta = 1$
Scaling in d	$d \in \{2, 4, 10\}$ at $K = 64, \Delta = 1$
Scaling in Δ	$\Delta \in \{0.5, 0.75, 1, 1.25, 1.5\}$ at $K = 64, d = 10$
Scaling in δ	$\delta \in \{0.2, 0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001\}$ at $K = 64, d = 10, \Delta = 1$

In the Convex-3D, Witness, and Arena settings, we first generate s Pareto models and then construct the dominated models by lowering the scores accordingly. The latent scores corresponding to the Pareto models are given by $\theta_i = 0.15\mathbf{1}_d + 0.75z_i$, where $z_i \overset{\text{iid}}{\sim} \text{Dirichlet}(\alpha\mathbf{1}_d)$, $\forall i \in [s]$; thus α controls their concentration on the trade-off simplex. For Convex-2D, the s Pareto anchors are deterministic and equally spaced on a decreasing line segment:

$$t_p = \frac{p-1}{s-1}, \qquad \theta_p = (0.20 + 0.65t_p,\ 0.85 - 0.65t_p), \qquad p \in [s].$$

In the Arena and Convex settings, each dominated arm $i \notin [s]$ is independently assigned an anchor $P_i \sim \text{Unif}([s])$ and its latent score is generated as

$$\theta_i = \theta_{P_i} - m_i \text{ where } m_{i,r} \overset{\text{iid}}{\sim} \text{Unif}[\ell, u] \quad \forall i \notin [s]. \tag{20}$$

For Convex-2D, $[\ell, u] = [0.03, 0.18]$. For the Witness setting, exactly q dominated arms are attached to each Pareto anchor $p \in [s]$, so $K = s(q+1)$. Candidate scores are generated according to (20) with $P_i = p$ and retained only when p is their unique Pareto dominator. In the Two-group-10 setting, we generate K_{low} low-group arms and K_{high} high-group arms, with $K = K_{\text{low}} + K_{\text{high}}$. Each coordinate is drawn independently from $\text{Unif}[0.20, 0.45]$ for the low-group arms and from $\text{Unif}[0.55, 0.75]$ for the high-group arms. Because every high-group coordinate exceeds every low-group coordinate, each low-group arm is strictly dominated by every high-group arm. The strict Pareto frontier therefore consists entirely of high-group arms, with its membership varying across realizations. The specific parameters used in each setting are reported in Tab. 2.

For Convex-3D and Arena-10, we generate 50 sets of latent model scores and simulate comparisons independently 10 times for each setting. The other benchmarks use 500 independently generated instances. For benchmarks with repeated observations of the same instance, standard errors are computed from the 50 instance-level averages.

Table 2: Parameters used in the synthetic benchmark experiments.

Setting	Generator	K	d	$\lvert\mathcal{P}\rvert$	Parameters
Convex-2D	Convex frontier	60	2	15	$s = 15$, margins $[0.03, 0.18]$
Convex-3D	Convex frontier	60	3	15	$s = 15$, simplex $\alpha = 1$, margins $[0.03, 0.18]$
Witness-4	Unique witness	40	4	8	$s = 8, q = 4$, margins $[0.04, 0.16]$, $\alpha = 0.7$
Witness-10	Unique witness	80	10	16	$s = 16, q = 4$, margins $[0.035, 0.14]$, $\alpha = 0.7$
Arena-4 small	Arena trade-off	32	4	8	$s = 8$, margins $[0.08, 0.25]$, $\alpha = 0.7$
Arena-4 medium	Arena trade-off	64	4	12	$s = 12$, margins $[0.08, 0.25]$, $\alpha = 0.7$
Arena-10 medium	Arena trade-off	64	10	16	$s = 16$, margins $[0.06, 0.20]$, $\alpha = 0.7$
Two-group-10	High-dimensional two-group	64	10	15.73	$K_{\text{low}} = 48, K_{\text{high}} = 16$

C.2 Implementation details for the baseline models

For *BT-MLE*, at each phase, we refit a centered ridge-regularized Bradley–Terry model per objective using all accumulated comparisons. For numerical stability, we use a ridge penalty coefficient of 10^{-8} and an optimization tolerance of 10^{-9}. If the optimizer fails to converge, any fitted score is nonfinite or exceeds 20 in absolute value, or the comparison graph is disconnected, we refit with coefficient 10^{-4}.

For *Borda plug-in*, let $N^{\text{pair}}_{ij,r}(t)$ denote the number of comparisons between models i and j on objective r up to query t, and let $W^{\text{on}}_{ij,r}(t)$ denote the number of these comparisons won by i. For $i < j$, we estimate the probability that i beats j by

$$\widetilde{p}^{(r)}_{ij}(t) := \frac{W^{\text{on}}_{ij,r}(t) + 1/2}{N^{\text{pair}}_{ij,r}(t) + 1}, \qquad \widetilde{p}^{(r)}_{ji}(t) := 1 - \widetilde{p}^{(r)}_{ij}(t).$$

The added counts prevent estimates of exactly zero or one. We estimate each model's Borda score on objective r by averaging its estimated win probabilities over all other models:

$$\widehat{b}_{i,r}(t) := \frac{1}{K-1} \sum_{j \neq i} \widetilde{p}^{(r)}_{ij}(t).$$

C.3 Additional results from the synthetic benchmark settings

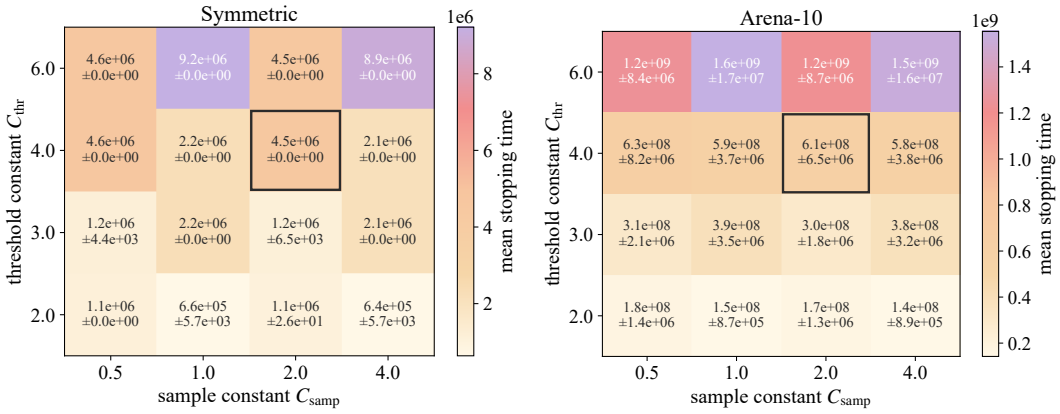
Tab. 3 presents the mean stopping time, the averaged ratio between the baselines stopping time to that of VB-EGE. We can see that VB-EGE has the lowest mean cost in all eight settings.

Table 3: Mean comparison counts for VB-EGE and mean comparison-count ratios for the baselines. Parentheses contain one standard error, computed from instance-level averages when multiple trials share the same latent instance.

Setting	VB mean τ (SE)	Focal/VB (SE)	MLE/VB (SE)	Plug-in/VB (SE)
Convex-2D	1.21×10^8 (9.57×10^5)	3.46 (0.10)	6.74 (0.21)	115 (3.50)
Convex-3D	4.46×10^9 (3.86×10^9)	9.85 (1.08)	21.2 (2.36)	341 (38.8)
Witness-4	2.28×10^8 (7.27×10^5)	1.26 (0.00)	1.80 (0.01)	29.8 (0.10)
Witness-10	1.60×10^9 (3.26×10^6)	1.81 (0.06)	2.98 (0.01)	49.1 (0.09)
Arena-4 small	3.83×10^{10} (3.30×10^{10})	9.46 (0.36)	10.1 (0.38)	170 (6.15)
Arena-4 medium	2.20×10^{10} (9.61×10^9)	14.8 (0.61)	32.7 (1.15)	581 (25.7)
Arena-10 medium	6.15×10^8 (5.87×10^6)	2.07 (0.02)	4.87 (0.05)	80.1 (0.80)
Two-group-10	4.48×10^{11} (3.69×10^{11})	17.9 (0.65)	38.7 (1.22)	649 (20.2)

C.4 Hyperparameter sensitivity analysis

We now examine sensitivity to the sampling and removal-threshold parameters by varying $C_{\text{samp}} \in \{0.5, 1, 2, 4\}$ and $C_{\text{thr}} \in \{2, 3, 4, 6\}$, while fixing $C_{\log} = 4$ and $\delta = 0.05$. We consider three settings: (1) the hard instance considered in the scaling studies with $K = 64$, $d = 10$, (2) the Arena-10 medium setting, and (3) the Witness-10 setting. We note that configurations other than the default are not covered by Thm. 3.1 as stated. Fig. 3 present heatmaps corresponding to the sensitivity analysis results. We see that smaller C_{thr} generally yields smaller stopping time.



(a) Symmetric $K = 64, d = 10$.

(b) Arena-10 medium.

(c) Witness-10.

Figure 3: Mean stopping-time with respect to difference choices of $(C_{\text{samp}}, C_{\text{thr}})$. Each cell prints mean $\pm$ one standard error. The black rectangle corresponds to the default choice.

C.5 Detailed fixed-confidence scaling results

Tab. 4 reports the values underlying panels (a)–(c) of Fig. 2a.

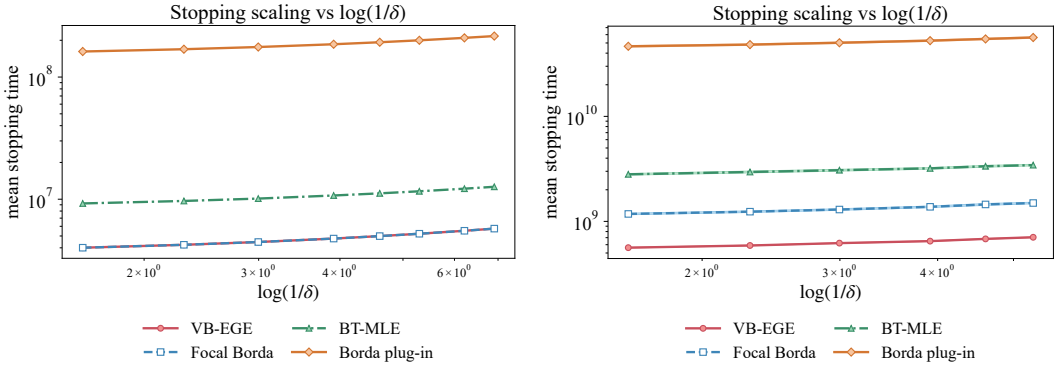
Table 4: Fixed-confidence scaling results. Parentheses contain one standard error for VB-EGE; remaining columns report each baseline's mean comparison count divided by that of VB-EGE.

Sweep	Value	VB mean τ (SE)	Focal/VB	MLE/VB	Plug-in/VB
K	16	3.63×10^5 (678)	1.02	0.51	9.03
	32	7.85×10^5 (411)	1.00	1.08	19.1
	64	1.66×10^6 (457)	1.00	2.29	40.1
	128	3.51×10^6 (301)	1.00	4.80	83.2
d	2	7.65×10^5 (1.63×10^3)	1.03	2.38	41.7
	4	1.66×10^6 (405)	1.00	2.29	40.1
	10	4.46×10^6 (0)	1.00	2.27	39.5
Δ	0.50	1.84×10^7 (2.78×10^3)	1.00	2.31	39.3
	0.75	4.46×10^6 (0)	1.00	2.27	39.5
	1.00	4.46×10^6 (0)	1.00	2.27	39.5
	1.25	1.82×10^6 (2.04×10^4)	2.45	1.28	23.4
	1.50	1.07×10^6 (212)	1.01	2.17	39.8

C.6 Confidence scaling and tail quantiles

We write $\bar{\tau}$ for the sample mean of stopping times across replications and $q_{0.95}(\tau)$ for their empirical 95th percentile.

The symmetric confidence sweep fixes $K = 64, d = 10, \Delta = 1$ and uses 500 replications at each $\delta \in \{0.2, 0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001\}$. The companion Arena-10 sweep uses $K = 64, d = 10, |\mathcal{P}| = 16$, margins $[0.06, 0.20]$, and 300 replications at each $\delta \in \{0.2, 0.1, 0.05, 0.02, 0.01, 0.005\}$. Within a replication, all δ values reuse the same latent instance, so confidence is not confounded with between-instance hardness.



(a) Symmetric $K = 64, d = 10$.

(b) Arena-10 medium.

Figure 4: Mean stopping time versus $\log(1/\delta)$. Shaded ribbons are mean $\pm$ one standard error.

Table 5: Confidence scaling on the symmetric $K = 64, d = 10$ task.

δ	VB mean τ (SE)	VB $q_{0.95}(\tau)$	Focal/VB	MLE/VB	Plug-in/VB
0.200	4.01×10^6 (0)	4.01×10^6	1.00	2.30	40.4
0.100	4.23×10^6 (0)	4.23×10^6	1.00	2.29	39.9
0.050	4.46×10^6 (0)	4.46×10^6	1.00	2.27	39.5
0.020	4.76×10^6 (0)	4.76×10^6	1.00	2.25	39.0
0.010	4.99×10^6 (0)	4.99×10^6	1.00	2.24	38.6
0.005	5.22×10^6 (0)	5.22×10^6	1.00	2.23	38.3
0.002	5.52×10^6 (0)	5.52×10^6	1.00	2.21	38.0
0.001	5.74×10^6 (0)	5.74×10^6	1.00	2.20	37.7

Table 6: Confidence scaling on the Arena-10 medium task.

δ	VB mean τ (SE)	VB $q_{0.95}(\tau)$	Focal/VB	MLE/VB	Plug-in/VB
0.200	5.62×10^8 (3.69×10^6)	6.57×10^8	2.09	4.99	82.7
0.100	5.89×10^8 (3.63×10^6)	6.91×10^8	2.10	5.01	81.8
0.050	6.21×10^8 (3.63×10^6)	7.10×10^8	2.09	4.95	80.9
0.020	6.49×10^8 (4.46×10^6)	7.69×10^8	2.12	4.93	81.1
0.010	6.81×10^8 (4.20×10^6)	7.86×10^8	2.13	4.92	80.2
0.005	7.06×10^8 (4.67×10^6)	8.20×10^8	2.12	4.87	79.9

The fits give $R^2 \approx 0.99997$ (symmetric) and 0.9978 (Arena-10) for $\bar{\tau} = a + b \log(1/\delta)$, with the same ordering at the 95th percentile. Thus stopping cost is empirically logarithmic in confidence; with 300–500 replications, these experiments cannot reliably estimate failure probabilities of order 10^{-3}.

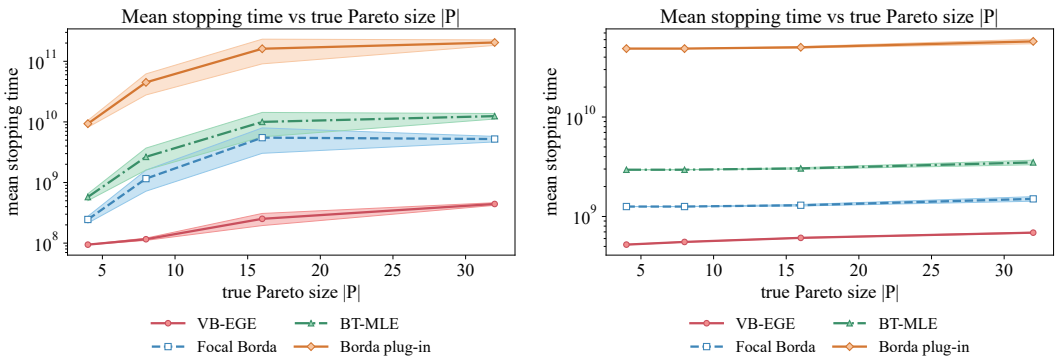
C.7 Pareto-set size sensitivity analysis

We hold $K = 64$ and the Arena-like generator family fixed while varying $|\mathcal{P}| \in \{4, 8, 16, 32\}$. Arena-4 uses margins $[0.08, 0.25]$, Arena-10 uses $[0.06, 0.20]$, and each parameter setting has 300 trials.

Table 7: Sensitivity to Pareto-set size. Parentheses contain one standard error for VB-EGE; other entries report each baseline's mean comparison count divided by that of VB-EGE.

d	$\lvert\mathcal{P}\rvert$	VB mean τ (SE)	Focal/VB	MLE/VB	Plug-in/VB
4	4	9.44×10^7 (7.02×10^5)	2.60	6.11	99.1
4	8	1.16×10^8 (4.91×10^6)	10.0	22.9	387
4	16	2.52×10^8 (5.70×10^7)	21.8	39.7	640
4	32	4.42×10^8 (2.22×10^7)	11.8	28.2	462
10	4	5.21×10^8 (2.79×10^6)	2.41	5.66	93.5
10	8	5.55×10^8 (3.11×10^6)	2.27	5.32	87.9
10	16	6.10×10^8 (4.21×10^6)	2.13	4.99	82.4
10	32	6.88×10^8 (6.87×10^6)	2.19	5.09	83.7

Stopping cost increases with $|\mathcal{P}|$, while every tested ratio of mean comparison counts (baseline/VB-EGE) remains above one; the adaptive advantage therefore persists across frontier sizes 4–32.



(a) Arena-4. (b) Arena-10.

Figure 5: Mean stopping time as the true Pareto set grows. Shaded ribbons are mean ± one standard error.

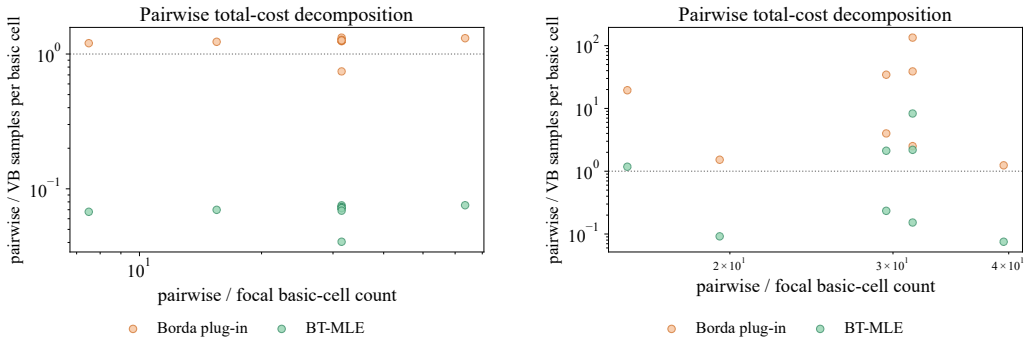
C.8 Sampling-mechanism diagnostics

The pairwise baselines sample $dK(K-1)/2$ pair–objective tuples, whereas VB-EGE samples Kd model–objective pairs. The ratio of these counts is $(K-1)/2$. Fig. 6 compares this ratio with the ratio of the average numbers of comparisons per tuple:

$$\frac{\tau_{\text{pair}}/[dK(K-1)/2]}{\tau_{\text{VB}}/(Kd)}.$$

Their product is the comparison-count ratio (pairwise baseline/VB-EGE).

This decomposition separates the number of sampled tuples from the average number of comparisons per tuple; it does not by itself imply an unavoidable sample-complexity penalty. Fig. 7 shows that VB-EGE collects more observations for models with smaller classification gaps.



(a) Scaling tasks. (b) Benchmark tasks.

Figure 6: Decomposition of comparison counts into the number of sampled tuples and the average number of comparisons per tuple.

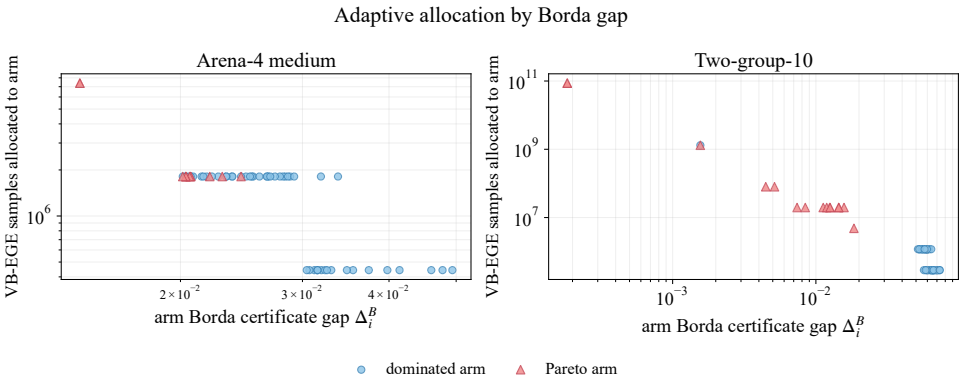


Figure 7: Total VB-EGE allocation to each model versus its Borda classification gap on representative Arena-4 medium and Two-group-10 instances. Smaller-gap models receive more samples.

D Additional real-data experiments details and results

This section provides details of preprocessing, sampling, and results for the three real-world Arena datasets. We also present the complete dataset-level sensitivity diagnostics for `tie` and `both_bad`, noting that these diagnostics are kept separate from the main comparison because changing the treatment of non-decisive labels changes the empirical target itself.

D.1 Datasets, preprocessing, and retained models

Table 8: Main realistic Arena data settings. Rows are retained pair–objective records; the Vision rows arise from 12,937 unique multi-label comparisons.

Dataset	Objectives d	Candidates	Retained K	Rows	$\lvert\mathcal{P}_{\text{emp}}\rvert$
Arena-140K	5	52	35	73,135	3
Search Arena 24K	9	13	8	4,305	5
VisionArena-Battle	8	17	12	23,458	3

Fixed dataset versions and definition of the objectives. All dataset revisions are fixed before preprocessing. Objective membership is determined exclusively from the released metadata; prompt and response text is not used to assign comparisons to objectives.

- **Arena Human Preference 140K:** `lmarena-ai/arena-human-preference-140k`.
  Revision: `6322995ab34d7c2693e3f47dd13fa5caa0789a74`. We construct five mutually exclusive prompt categories: `general`, `code`, `math`, `creative_writing`, and `instruction_following`. The four specialized categories are identified using the released metadata:
  - `is_code` for `code`,
  - `category_tag.math_v0.1.math` for `math`,
  - `category_tag.creative_writing_v0.1.creative_writing` for `creative_writing`,
  - `category_tag.if_v0.1.if` for `instruction_following`.

  The `general` category is a residual category containing comparisons not assigned to any of these four specialized categories.
  Because the released indicators may overlap, we apply the prespecified deterministic priority

  `code` > `math` > `creative_writing` > `instruction_following` > `general`.

Each comparison is assigned to the first applicable category in this ordering. For example, a comparison marked as both code and math is assigned to `code`, while a comparison with no active specialized indicator is assigned to `general`. Thus, every comparison assigned to a category contributes to exactly one Arena objective, preventing the same human judgment from being reused across multiple Arena coordinates. The five-category construction, priority ordering, and residual definition of `general` are our preprocessing choices rather than an official five-way partition released with the dataset.

- **Search Arena 24K:** `lmarena-ai/search-arena-24k`.
  Revision: `fac8dcf86146c8773ef020095c5694c9b2bc98d7`. We use the nine released values of the categorical `primary_intent` field as objectives: Factual Lookup, Info Synthesis, Recommendation, Analysis, Explanation, Guidance, Creative Generation, Other, and Text Processing. Since `primary_intent` is single-valued, each retained comparison contributes to at most one objective. Comparisons with a missing or unsupported intent are not assigned to a pair–objective cell; no additional fallback category is constructed.
- **VisionArena-Battle:** `lmarena-ai/VisionArena-Battle`.
  Revision: `a41d58f15caa3e2d9685774153d2bc815b943d96`. We use the eight released Boolean category indicators directly: `captioning`, `creative_writing`, `diagram`, `entity_recognition`, `homework`, `humor`, `is_code`, and `ocr`. These annotations are multi-label by construction. We therefore preserve their released semantics rather than imposing an artificial priority ordering: a comparison contributes once to every pair–objective cell whose corresponding indicator is true. Consequently, the number of pair–objective assignments may exceed the number of unique source comparisons; these repeated assignments represent overlapping category membership, not additional independent human votes.

Across datasets, the common principle is to preserve the semantics of the released annotation schema whenever it directly defines the objectives: single-valued categorical annotations produce mutually exclusive assignments, whereas released Boolean multi-label annotations remain non-exclusive. Arena is the only exception requiring a separate rule for defining the objectives; its prespecified priority rule provides disjoint prompt categories and makes this additional preprocessing choice explicit.

The main Arena experiments retain only comparisons in which model a or model b wins. Arena removes 21,528 `tie` and 15,749 `both_bad` records. Search removes 2,282 `tie`, 1,442 `both_bad` records, 10,751 missing outcomes, and 1,285 comparisons whose two model names refer to the same model after name normalization. Vision removes 5,216 `tie` and 5,905 `both_bad` records; another 2,762 decisive records match none of the eight official category flags. These outcome exclusions occur before candidate eligibility and model selection.

Table 9: Preprocessing counts for the three Arena datasets. "Assigned comparisons" counts source comparisons assigned to at least one official objective; "Objective records" counts the resulting pair–objective records. The last column uses the models retained at $q = 1$.

Dataset	Source	Decisive	Assigned comparisons	Objective records	Candidates	K	Retained rows
Arena-140K	135,609	98,332	98,332	98,332	52	35	73,135
Search Arena 24K	24,069	8,309	8,309	8,309	13	8	4,305
VisionArena-Battle	29,849	18,728	15,966	28,925	17	12	23,458

The retained Vision rows correspond to 12,937 unique source comparisons. Restriction to the retained models excludes 25,197, 4,004, and 5,467 pair–objective records already assigned to objectives from Arena, Search, and Vision, respectively; the Vision exclusion represents 3,029 unique source comparisons. These counts are downstream of the outcome filter and therefore must not be added to the `tie` and `both_bad` counts as if they were disjoint removals from the raw source table.

Selection of the retained models. At threshold q, two candidate models are adjacent when their unordered pair has at least q decisive observations in every configured objective. We choose an exact maximum clique. Ties are broken first by the total retained within-clique pair–objective support and then by the lexicographically smallest sorted model tuple. The main experiments all use $q = 1$, producing the fixed sets of retained models in Tab. 8. Their direct empirical strict frontiers are:

Table 10: Full-dimensional strict empirical Pareto sets in the three main sets of models retained at $q = 1$.

Dataset	Pareto members
Arena-35	`chatgpt-4o-latest-20250326`; `gemini-2.5-pro`; `grok-3-preview-02-24`
SearchArena-8	`gemini-2.5-pro-exp-03-25-grounding`; `sonar`; `sonar-pro-high`; `sonar-reasoning`; `sonar-reasoning-pro-high`
VisionArena-12	`claude-3-5-sonnet-20240620`; `gemini-1.5-pro-exp-0801`; `gpt-4o-2024-05-13`

D.2 Full Pareto-set results and method comparisons

We organize the plots by diagnostic and present Arena-35, SearchArena-8, and VisionArena-12 in that order within each diagnostic. In the full-dimensional profile plots, Pareto membership is computed using all five, nine, or eight objectives, respectively. In the two- and three-dimensional plots, it is recomputed using only the displayed objectives, so these frontiers may differ from the full-dimensional target.

D.2.1 Full-dimensional empirical Borda profiles

In Figs. 8 to 10, colored lines are strict-frontier members and faint gray lines are dominated members of the same retained model set. These plots retain the original Borda scale and all retained models, so they show both score magnitudes and the frontier's separation from dominated alternatives.

Across all three datasets, the frontier profiles cross repeatedly across objectives, showing that Pareto membership is driven by complementary strengths rather than one uniformly best model.

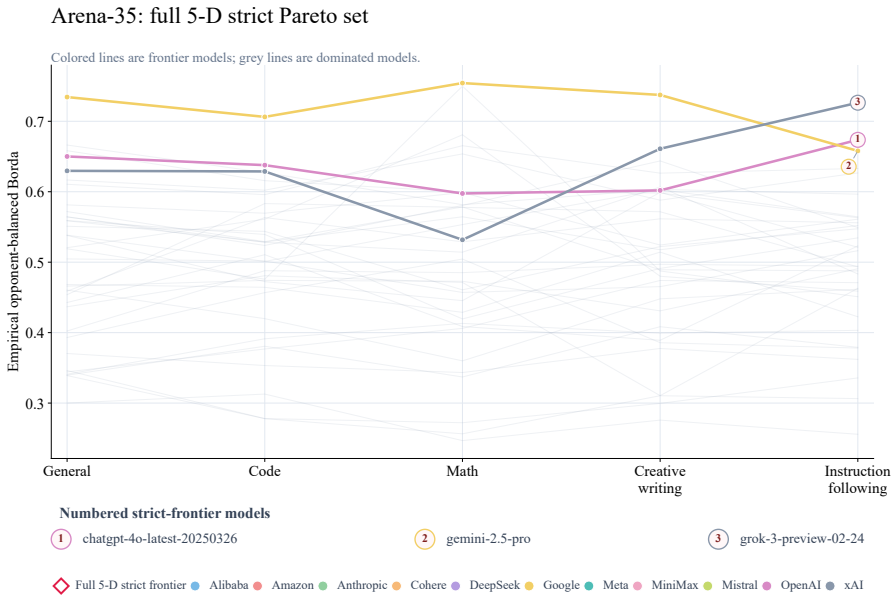


Figure 8: Arena-35: empirical Borda profiles on all five objectives.

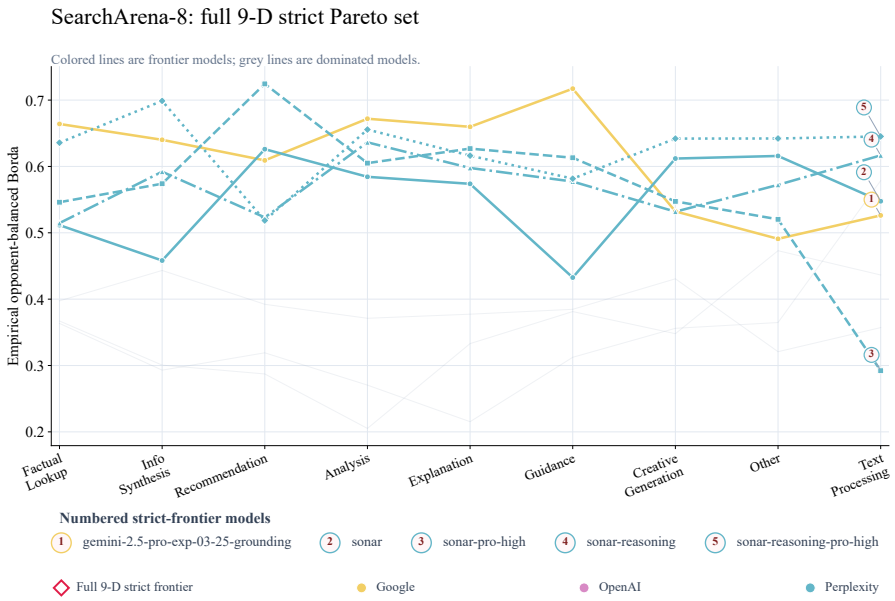


Figure 9: SearchArena-8: empirical Borda profiles on all nine objectives.

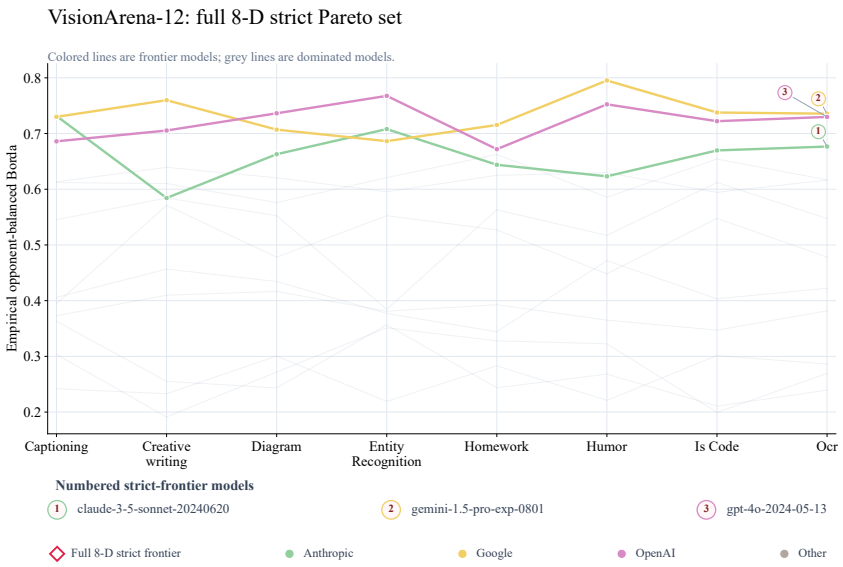
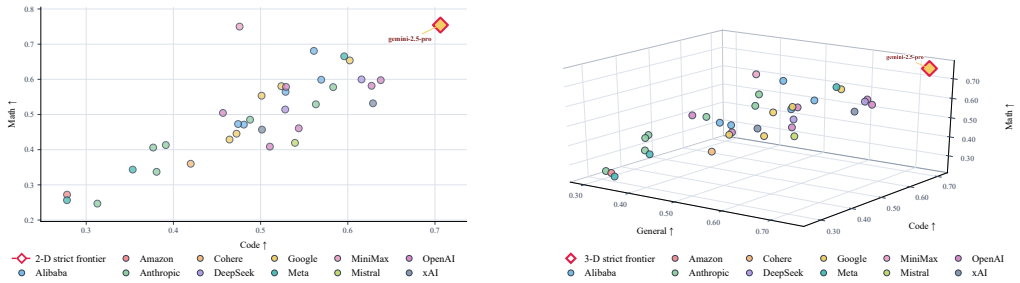


Figure 10: VisionArena-12: empirical Borda profiles on all eight objectives.

D.2.2 Low-dimensional explanatory projections

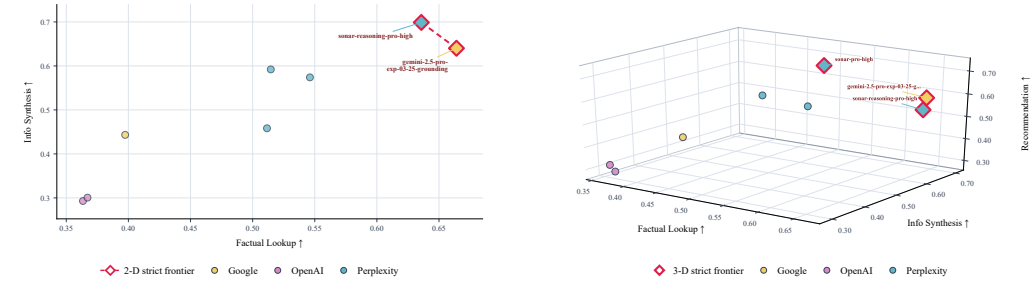
Figs. 11 to 13 pair representative two- and three-dimensional projections for each dataset. Pareto membership is recomputed in each projected space, so these views illustrate selected trade-offs but need not reproduce the full-dimensional frontier. Results for all objective pairs are available in the project repository.



(a) 2D: Code versus Math.

(b) 3D: General, Code, and Math.

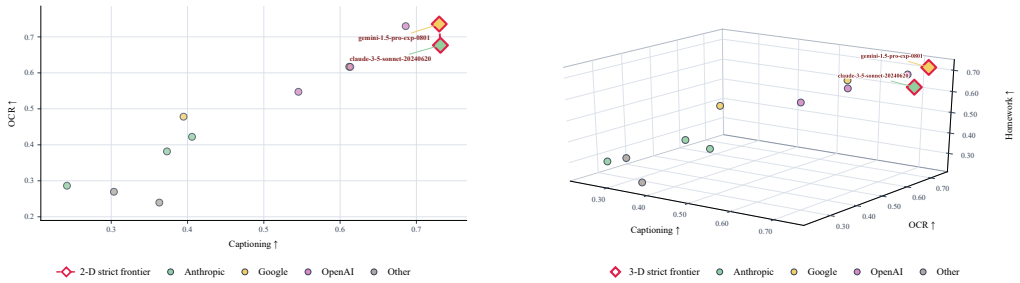
Figure 11: Arena-35 explanatory projections.



(a) 2D: Factual Lookup versus Info Synthesis.

(b) 3D: Factual Lookup, Info Synthesis, and Recommendation.

Figure 12: SearchArena-8 explanatory projections.



(a) 2D: Captioning versus OCR.

(b) 3D: Captioning, OCR, and Homework.

Figure 13: VisionArena-12 explanatory projections.

D.2.3 Comparison of the four methods on the same Arena datasets

Table 11: Comparison of the four methods on the three Arena datasets. Entries are mean stopping times in millions of comparisons, with each method's mean divided by the VB-EGE mean in parentheses. Every method stops and recovers the exact frontier in all 100 trials on each dataset.

Method	Arena-35	SearchArena-8	VisionArena-12
VB-EGE	86.82 (1.00×)	124.34 (1.00×)	8.80 (1.00×)
Uniform focal Borda	1,435.70 (16.54×)	492.74 (3.96×)	40.39 (4.59×)
Uniform pairwise BT-MLE	1,586.26 (18.27×)	149.06 (1.20×)	14.77 (1.68×)
Uniform pairwise Borda plug-in	26,164.10 (301.36×)	1,890.10 (15.20×)	251.62 (28.60×)

We can find that VB-EGE has the lowest mean cost on every dataset. The closest competitor is focal Borda on Arena-35 and BT-MLE on SearchArena-8 and VisionArena-12, while the Borda plug-in is consistently the most expensive method.

D.3 Sampling from logged Arena comparisons

We evaluate VB-EGE by sampling with replacement from logged Arena comparisons, without fitting a parametric comparison model. Let $\mathcal{D}_{\mathrm{Arena}}$ denote the fixed binary comparison log used in the experiment. All statements up to the final population-level lemma are conditional on this observed log: once $\mathcal{D}_{\mathrm{Arena}}$ is fixed, the only randomness comes from sampling opponents and logged outcomes and the internal randomization of the identification algorithm. We use the subscript samp on probabilities and expectations to denote this randomness.

Logged comparisons grouped by model pair and objective. We fix the set of retained models so that every model pair has at least one logged binary comparison for every objective. For objective $r \in [d]$, let $N_{ij}^{(r)}$ be the number of logged binary comparisons between models i and j. We require

$$N_{ij}^{(r)} \geq 1, \qquad \forall\, 1 \leq i < j \leq K, \quad r \in [d]. \tag{21}$$

We call the records for one model pair and one objective a *pair–objective cell*. For $i < j$, write its outcomes as

$$\mathcal{D}_{ij}^{(r)} := \left(Y_{ij,s}^{(r)}\right)_{s=1}^{N_{ij}^{(r)}},$$

where $Y_{ij,s}^{(r)} = 1$ means that model i defeats model j in the s-th logged comparison for objective r. For the reverse orientation, define

$$Y_{ji,s}^{(r)} := 1 - Y_{ij,s}^{(r)}, \qquad N_{ji}^{(r)} := N_{ij}^{(r)}.$$

Write $\mathcal{D}_{ji}^{(r)} := (Y_{ji,s}^{(r)})_{s=1}^{N_{ji}^{(r)}}$. Each cell is an indexed sequence of records, so repeated binary outcomes retain their multiplicities. Uniform sampling from a cell always means uniform sampling of its record index. The empirical pairwise win probability is

$$\widehat{p}_{ij}^{(r)} := \frac{1}{N_{ij}^{(r)}} \sum_{s=1}^{N_{ij}^{(r)}} Y_{ij,s}^{(r)}, \qquad \widehat{p}_{ji}^{(r)} := 1 - \widehat{p}_{ij}^{(r)}.$$

The empirical uniform-opponent vector-Borda coordinate of model i on objective r is

$$b_{i,r}^{\mathrm{emp}} := \frac{1}{K-1} \sum_{j \neq i} \widehat{p}_{ij}^{(r)}. \tag{22}$$

Let

$$b_i^{\mathrm{emp}} := \left(b_{i,1}^{\mathrm{emp}}, \ldots, b_{i,d}^{\mathrm{emp}}\right), \qquad b^{\mathrm{emp}} := \left(b_1^{\mathrm{emp}}, \ldots, b_K^{\mathrm{emp}}\right).$$

The empirical ground truth is the strict Pareto frontier of the empirical Borda vectors:

$$\mathcal{P}_{\mathrm{emp}} := \mathcal{P}\left(b^{\mathrm{emp}}\right).$$

This target is defined directly from the logged pair–objective cells. In particular, the empirical pairwise probability arrays $\{\widehat{p}_{ij}^{(r)}\}_{i,j,r}$ need not admit a global Bradley–Terry representation.

Sampling an opponent and a logged outcome. Whenever VB-EGE requests a focal-model sample for model i and objective r, we first draw an opponent uniformly from all other retained models, including models previously removed from the active set, and then draw one logged outcome uniformly with replacement from the corresponding pair–objective cell:

$$J \sim \mathrm{Unif}([K] \setminus \{i\}),$$
$$S \mid J = j \sim \mathrm{Unif}\left(\{1, \dots, N_{ij}^{(r)}\}\right). \tag{23}$$

The resulting binary observation is

$$Z := Y_{iJ,S}^{(r)}. \tag{24}$$

All opponent draws and within-cell draws are performed independently with replacement and independently of the algorithmic randomization.

Lemma D.1 (Conditional Bernoulli representation of sampling from logged comparisons)**.** *Fix the logged dataset $\mathcal{D}_{\mathrm{Arena}}$. For every focal model $i \in [K]$ and objective $r \in [d]$, an observation generated by (23)–(24) satisfies*

$$\mathbb{E}_{\mathrm{samp}}\left[Z \mid \mathcal{D}_{\mathrm{Arena}}, i, r\right] = b_{i,r}^{\mathrm{emp}}, \tag{25}$$

and

$$Z \mid \mathcal{D}_{\mathrm{Arena}}, i, r \sim \mathrm{Bernoulli}\left(b_{i,r}^{\mathrm{emp}}\right). \tag{26}$$

Moreover, the sampling randomization can be coupled to mutually independent streams, one for each model–objective pair, such that

$$Z_{i,r,1}, Z_{i,r,2}, \dots \mid \mathcal{D}_{\mathrm{Arena}} \overset{\mathrm{iid}}{\sim} \mathrm{Bernoulli}\left(b_{i,r}^{\mathrm{emp}}\right), \qquad i \in [K],\ r \in [d]. \tag{27}$$

Consequently, for every fixed $n \geq 1$,

$$\overline{Z}_{i,r}(n) := \frac{1}{n} \sum_{s=1}^{n} Z_{i,r,s}$$

is the mean of the first n stream observations; in particular, $\widehat{b}_{i,r}^{(m)} = \overline{Z}_{i,r}(n_m)$ at phase m. For fixed n, this sample mean is conditionally unbiased:

$$\mathbb{E}_{\mathrm{samp}}\left[\overline{Z}_{i,r}(n) \mid \mathcal{D}_{\mathrm{Arena}}\right] = b_{i,r}^{\mathrm{emp}}.$$

Proof. Throughout the proof, condition on the fixed dataset $\mathcal{D}_{\mathrm{Arena}}$. The contents of every logged cell are then deterministic, and the only remaining randomness comes from sampling the opponent and the logged outcome.

Fix a focal model i and objective r. Conditional on drawing opponent $J = j$, the sampling procedure draws one record uniformly from $\mathcal{D}_{ij}^{(r)}$. Therefore

$$\begin{aligned}\mathbb{E}_{\mathrm{samp}}\left[Z \mid \mathcal{D}_{\mathrm{Arena}}, i, r, J = j\right] &= \frac{1}{N_{ij}^{(r)}} \sum_{s=1}^{N_{ij}^{(r)}} Y_{ij,s}^{(r)} \\ &= \widehat{p}_{ij}^{(r)}.\end{aligned}$$

Averaging over the uniform opponent draw gives

$$\begin{aligned}\mathbb{E}_{\mathrm{samp}}\left[Z \mid \mathcal{D}_{\mathrm{Arena}}, i, r\right] &= \sum_{j \neq i} \mathbb{P}_{\mathrm{samp}}(J = j) \mathbb{E}_{\mathrm{samp}}\left[Z \mid \mathcal{D}_{\mathrm{Arena}}, i, r, J = j\right] \\ &= \frac{1}{K-1} \sum_{j \neq i} \widehat{p}_{ij}^{(r)} \\ &= b_{i,r}^{\mathrm{emp}}.\end{aligned}$$

This proves (25).

Because $Z \in \{0, 1\}$, its conditional distribution is determined by its conditional success probability, which proves (26). Independent sampling steps use fresh opponent and within-cell indices, each

drawn with replacement. Conditional on the fixed log, these draws therefore generate mutually independent streams across (i, r), with each stream iid Bernoulli with mean $b_{i,r}^{\mathrm{emp}}$. This proves (27). Linearity of conditional expectation then gives

$$\mathbb{E}_{\mathrm{samp}}\left[\overline{Z}_{i,r}(n) \mid \mathcal{D}_{\mathrm{Arena}}\right] = \frac{1}{n}\sum_{s=1}^{n} \mathbb{E}_{\mathrm{samp}}\left[Z_{i,r,s} \mid \mathcal{D}_{\mathrm{Arena}}\right] = b_{i,r}^{\mathrm{emp}}.$$

The stream representation remains valid when the focal pairs are selected adaptively. Equivalently, one may generate all streams in advance and reveal the next unused element of stream (i, r) whenever the algorithm queries that pair. Thus adaptivity changes only which prefixes are revealed, not their conditional distributions. □

Corollary D.2 (Recovery guarantee conditional on the logged dataset)**.** *Fix a logged dataset $\mathcal{D}_{\mathrm{Arena}}$ satisfying (21). Define*

$$\Delta_i^{\mathrm{emp}} := \Delta_i\left(b^{\mathrm{emp}}\right), \qquad \Delta_{\min}^{\mathrm{emp}} := \min_{i\in[K]} \Delta_i^{\mathrm{emp}}.$$

Assume $\Delta_{\min}^{\mathrm{emp}} > 0$, and run Alg. 1 with the common configuration $C = (2, 4, 4)$ and the sampling procedure defined above. For each model i, let

$$M_i^{\mathrm{emp}} := 1 + \left\lceil \log_2 \frac{4}{\Delta_i^{\mathrm{emp}}} \right\rceil, \qquad L_i^{\mathrm{emp}}(\delta) := \log \frac{4Kd(M_i^{\mathrm{emp}})^2}{\delta}.$$

Then, conditional on the fixed dataset,

$$\mathbb{P}_{\mathrm{samp}}\left(\widehat{\mathcal{P}} = \mathcal{P}_{\mathrm{emp}}, \tau \le dK + 128d \sum_{i=1}^{K} \frac{L_i^{\mathrm{emp}}(\delta)}{(\Delta_i^{\mathrm{emp}})^2} \,\middle|\, \mathcal{D}_{\mathrm{Arena}}\right) \ge 1 - \delta.$$

Proof. Conditional on $\mathcal{D}_{\mathrm{Arena}}$, the vector b^{emp} is deterministic. By Lemma D.1, the sampling procedure supplies mutually independent Bernoulli streams with mean array b^{emp}, independently of the algorithmic randomization. Thus the conditional sampling law is exactly the vector-Bernoulli law in Prop. B.4 with $b = b^{\mathrm{emp}}$. Applying Prop. B.4 and using

$$\mathcal{P}(b^{\mathrm{emp}}) = \mathcal{P}_{\mathrm{emp}}$$

proves the claim. □

The Bernoulli representation in Lemma D.1 holds for any choice of the algorithmic constants. Under the common configuration $C = (2, 4, 4)$, Prop. B.4 yields the conditional finite-sample guarantee in Cor. D.2.

Why opponent-first sampling is necessary. The opponent must be sampled uniformly before a logged outcome is sampled from the selected cell. If instead one sampled a record uniformly by choosing an index from the pooled record-index set

$$\left\{(j, s) : j \in [K] \setminus \{i\},\ 1 \le s \le N_{ij}^{(r)}\right\},$$

and returning $Y_{ij,s}^{(r)}$, then opponent j would be selected with probability proportional to $N_{ij}^{(r)}$. The conditional mean of the sampled outcome would be

$$\frac{\sum_{j\neq i} N_{ij}^{(r)} \widehat{p}_{ij}^{(r)}}{\sum_{j\neq i} N_{ij}^{(r)}},$$

which is generally different from the uniform-opponent Borda coordinate $b_{i,r}^{\mathrm{emp}}$ in (22).

Interpretation of the comparison count. Each sampling step draws one binary outcome with replacement from the fixed historical log and counts as one comparison. Thus the stopping time counts all sampled comparisons, including repeated draws of the same historical record. All methods use the same retained models and logged outcomes and are evaluated against the same empirical ground truth $\mathcal{P}_{\mathrm{emp}}$.

Optional population-level interpretation. The preceding results condition directly on the observed Arena log. A population-level score interpretation requires an additional assumption on how the logged pair–objective cells were generated. Let

$$\mathbf{N} := \left\{ N_{ij}^{(r)} : 1 \le i < j \le K,\ r \in [d] \right\}$$

denote the array of cell counts.

Lemma D.3 (Population-level Borda unbiasedness under cell-wise iid logging)**.** *Suppose that, for every $1 \le i < j \le K$ and $r \in [d]$, conditional on $\mathbf{N}$, the logged outcomes in cell (i, j, r) satisfy*

$$Y_{ij,1}^{(r)}, \ldots, Y_{ij,N_{ij}^{(r)}}^{(r)} \mid \mathbf{N} \overset{\text{iid}}{\sim} \text{Bernoulli}\left(p_{ij}^{(r)}\right). \tag{28}$$

No independence across distinct pair–objective cells is needed for the mean identities below. Define

$$p_{ji}^{(r)} := 1 - p_{ij}^{(r)}$$

and the population uniform-opponent Borda coordinate

$$b_{i,r}^{\text{pop}} := \frac{1}{K-1} \sum_{j \ne i} p_{ij}^{(r)}.$$

Then, for every ordered pair $i \ne j$ and objective r,

$$\mathbb{E}_{\mathcal{D}}\left[\widehat{p}_{ij}^{(r)} \mid \mathbf{N}\right] = p_{ij}^{(r)},$$

and

$$\mathbb{E}_{\mathcal{D}}\left[b_{i,r}^{\text{emp}} \mid \mathbf{N}\right] = b_{i,r}^{\text{pop}}.$$

Moreover, for a single sampled observation and every fixed $n \ge 1$,

$$Z \mid \mathbf{N}, i, r \sim \text{Bernoulli}\left(b_{i,r}^{\text{pop}}\right), \tag{29}$$

and

$$\mathbb{E}_{\mathcal{D},\text{samp}}\left[\overline{Z}_{i,r}(n) \mid \mathbf{N}\right] = b_{i,r}^{\text{pop}}.$$

Proof. For $i < j$, the cell-wise assumption (28) gives

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}}\left[\widehat{p}_{ij}^{(r)} \mid \mathbf{N}\right] &= \frac{1}{N_{ij}^{(r)}} \sum_{s=1}^{N_{ij}^{(r)}} \mathbb{E}_{\mathcal{D}}\left[Y_{ij,s}^{(r)} \mid \mathbf{N}\right] \\
&= p_{ij}^{(r)}.
\end{aligned}$$

For the reverse orientation, $\widehat{p}_{ji}^{(r)} = 1 - \widehat{p}_{ij}^{(r)}$ and $p_{ji}^{(r)} = 1 - p_{ij}^{(r)}$, so the same identity holds for every ordered pair $i \ne j$.

Averaging over opponents yields

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}}\left[b_{i,r}^{\text{emp}} \mid \mathbf{N}\right] &= \frac{1}{K-1} \sum_{j \ne i} \mathbb{E}_{\mathcal{D}}\left[\widehat{p}_{ij}^{(r)} \mid \mathbf{N}\right] \\
&= \frac{1}{K-1} \sum_{j \ne i} p_{ij}^{(r)} \\
&= b_{i,r}^{\text{pop}}.
\end{aligned}$$

By Lemma D.1 and the tower property,

$$\begin{aligned}
\mathbb{P}_{\mathcal{D},\text{samp}}\left(Z = 1 \mid \mathbf{N}, i, r\right) &= \mathbb{E}_{\mathcal{D}}\left[\mathbb{P}_{\text{samp}}\left(Z = 1 \mid \mathcal{D}_{\text{Arena}}, i, r\right) | \mathbf{N}\right] \\
&= \mathbb{E}_{\mathcal{D}}\left[b_{i,r}^{\text{emp}} \mid \mathbf{N}\right] \\
&= b_{i,r}^{\text{pop}}.
\end{aligned}$$

Since Z is binary, this proves (29). Applying the tower property again gives

$$\begin{aligned}
\mathbb{E}_{\mathcal{D},\text{samp}}\left[\overline{Z}_{i,r}(n) \mid \mathbf{N}\right] &= \mathbb{E}_{\mathcal{D}}\left[\mathbb{E}_{\text{samp}}\left[\overline{Z}_{i,r}(n) \mid \mathcal{D}_{\text{Arena}}\right] \middle| \mathbf{N}\right] \\
&= \mathbb{E}_{\mathcal{D}}\left[b_{i,r}^{\text{emp}} \mid \mathbf{N}\right] \\
&= b_{i,r}^{\text{pop}}.
\end{aligned}$$

□

The population-level lemma is not used in the recovery guarantee conditional on the logged dataset. It requires the additional representative cell-wise logging assumption (28). After integrating over the random log, repeated sampled observations need not remain independent because they share the same realized empirical cell means; only the conditional iid statement in Lemma D.1 is used by the algorithmic analysis.

D.4 SENSITIVITY TO TIE AND BOTH_BAD HANDLING

The main experiments use binary feedback and therefore discard non-decisive labels. This section evaluates whether the empirical Pareto target is sensitive to two alternative treatments of the released `tie` and `both_bad` outcomes. The VB-EGE algorithm itself is unchanged throughout. We compare three target constructions:

D = decisive-only, H = direct merged half-credit, F = fitted Davidson half-credit.

For $X \in \{D, H, F\}$, write $b^X = (b_1^X, \dots, b_K^X)$, with $b_i^X = (b_{i,1}^X, \dots, b_{i,d}^X)$, for the fixed Borda vectors constructed below. Define

$$b^D := b^{\mathrm{emp}}, \qquad \mathcal{P}_X := \mathcal{P}(b^X), \qquad \Delta_{\min}^X := \min_{i \in [K]} \Delta_i(b^X).$$

Both alternatives start from all four released outcomes while holding the set of models retained at $q = 1$, assignment of comparisons to objectives, and uniform-opponent aggregation fixed. Keeping the retained models fixed prevents outcome handling from being confounded with selection of the models.

For an unordered pair $\{i, j\}$ and objective r, let $W_{ij,r}$ and $W_{ji,r}$ denote the two directed win counts, and let $T_{ij,r}$ and $B_{ij,r}$ denote the `tie` and `both_bad` counts after assigning comparisons to objectives as in the main experiments. We write

$$M_{ij,r} := T_{ij,r} + B_{ij,r}$$

for the merged non-directional count. The direct construction H changes only the reward assigned to $M_{ij,r}$ and recomputes the target nonparametrically. The fitted construction F additionally fits a three-outcome Davidson model to $(W_{ij,r}, W_{ji,r}, M_{ij,r})$, following the factored tie construction of Ameli et al. (2025), before supplying the resulting target to VB-EGE.

The merge $M_{ij,r} = T_{ij,r} + B_{ij,r}$ is an explicit sensitivity choice needed to interface the four released labels with a three-outcome Davidson likelihood. It does not assert that a `tie` judgment and a `both_bad` judgment have the same semantic meaning.

Table 12: Outcome counts on the fixed sets of models retained at $q = 1$ before non-decisive labels are merged. Each record represents one comparison assigned to one objective. Arena-35 and SearchArena-8 assign each comparison to a single objective, whereas VisionArena-12 is multi-label; consequently, the Vision counts are objective assignments rather than unique source comparisons. The last column is the fraction of these records labeled `tie` or `both_bad`.

Dataset	Decisive	`tie`	`both_bad`	Non-decisive fraction
Arena-35	73,135	16,093	11,594	27.5%
SearchArena-8	4,305	1,148	676	29.8%
VisionArena-12	23,458	6,346	8,031	38.0%

D.4.1 DIRECT MERGED HALF-CREDIT RECODING

The empirical win rate in the main experiment uses only comparisons with a winner,

$$\widehat{p}_{ij,r}^{D} = \frac{W_{ij,r}}{W_{ij,r} + W_{ji,r}} = \widehat{p}_{ij}^{(r)}.$$

The direct sensitivity keeps each decisive win coded as $(1, 0)$ and assigns $(1/2, 1/2)$ to every `tie` and `both_bad` observation. Its directed cell mean is therefore

$$\widehat{p}_{ij,r}^{H} = \frac{W_{ij,r} + \frac{1}{2} M_{ij,r}}{W_{ij,r} + W_{ji,r} + M_{ij,r}}, \qquad \widehat{p}_{ji,r}^{H} = 1 - \widehat{p}_{ij,r}^{H}.$$

No pooling or parametric fitting is used: every pair–objective cell is estimated from its own logged counts.

Target and comparison. For each model and objective, we recompute

$$b_{i,r}^{H} = \frac{1}{K-1} \sum_{j \neq i} \widehat{p}_{ij,r}^{H}$$

and obtain $\mathcal{P}_H$. We compare this exact half-credit target with the decisive-only target using membership, Jaccard overlap, model-wise score changes, and objective-wise Spearman rank correlations. We write

$$J(X,Y) := \frac{|\mathcal{P}_X \cap \mathcal{P}_Y|}{|\mathcal{P}_X \cup \mathcal{P}_Y|}, \qquad X, Y \in \{D, H, F\},$$

and let $\rho_{\min}(X,Y)$ denote the minimum, over objectives, of the Spearman correlation between the corresponding model-score vectors. This first analysis is a target diagnostic only: it does not rerun VB-EGE or report a new stopping cost.

Results. Relative to the decisive-only target D, direct half-credit recoding H changes strict Pareto membership only for Arena-35. It adds `minimax-m1`, increasing the frontier size from three to four and giving $J(D,H) = 0.75$. SearchArena-8 and VisionArena-12 retain their decisive-only frontiers, with $J(D,H) = 1$ for both datasets. The minimum objective-wise Spearman correlations between D and H are 0.993, 0.929, and 0.958 for Arena-35, SearchArena-8, and VisionArena-12, respectively. The complete comparison of both tie-aware constructions with the decisive-only baseline is reported in Tab. 14.

Dataset-level Pareto profiles. The following figures hold the diagnostic fixed and report the three datasets in the same order used throughout this appendix.

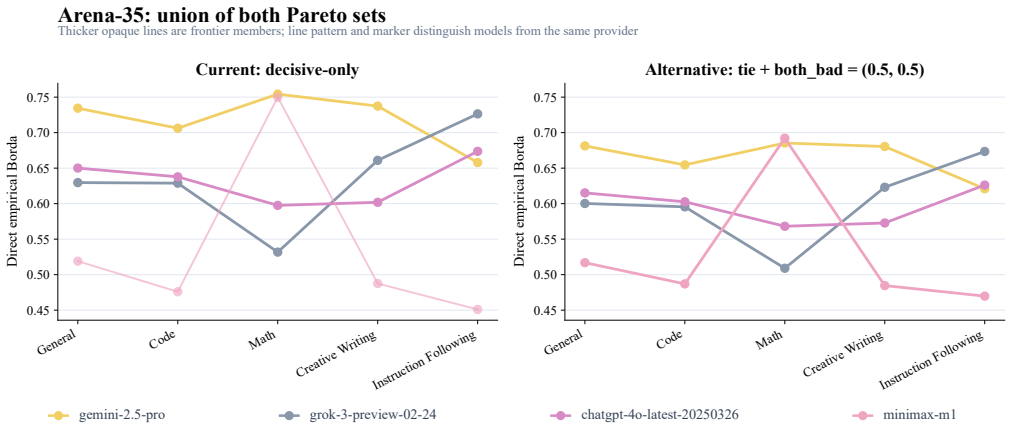


Figure 14: Arena-35: direct outcome-coding sensitivity. The dashed `minimax-m1` profile becomes Pareto under half-credit coding.

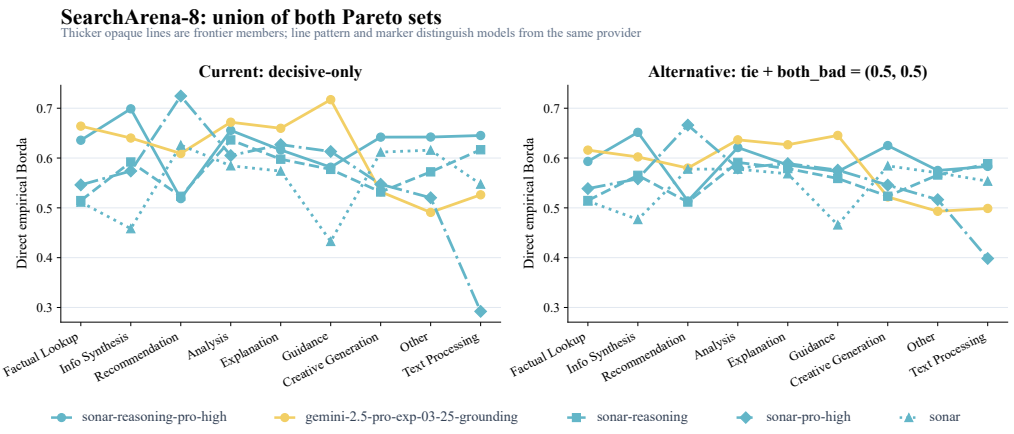


Figure 15: SearchArena-8: direct outcome-coding sensitivity; frontier membership is unchanged.

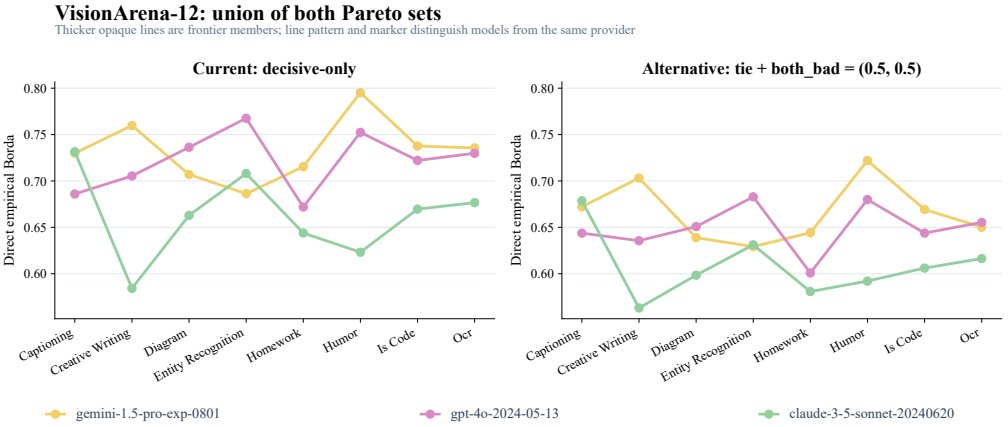


Figure 16: VisionArena-12: direct outcome-coding sensitivity; frontier membership is unchanged.

D.4.2 Factored-Davidson tie-rank selection

Three-outcome likelihood. The second analysis again merges `tie` and `both_bad`, but now uses the full count triple $(W_{ij,r}, W_{ji,r}, M_{ij,r})$ to fit a generalized Davidson model (Davidson, 1970) independently for every objective r, following the factored tie construction of Ameli et al. (2025). For the canonical ordering $i < j$, let $s_{i,r}$ be a model score and let $\eta_{ij,r}$ be the symmetric pair-specific log tie weight. The probabilities of "i wins," "j wins," and "merged tie" are

$$\left(\pi^{W}_{ij,r},\, \pi^{L}_{ij,r},\, \pi^{M}_{ij,r}\right) = \mathrm{softmax}\left(\frac{s_{i,r} - s_{j,r}}{2},\, \frac{s_{j,r} - s_{i,r}}{2},\, \eta_{ij,r}\right), \qquad \sum_{i=1}^{K} s_{i,r} = 0.$$

Thus the two directional outcomes determine relative strength, whereas the third outcome determines the propensity of that pair to receive either non-decisive label. Parameters minimize the three-category multinomial negative log likelihood obtained by weighting the three log probabilities in Sec. D.4.2 by $(W_{ij,r}, W_{ji,r}, M_{ij,r})$ and summing over $i < j$.

Factored tie weight. When $k_{\text{tie}} = 0$, all pairs in objective r share one scalar tie logit, $\eta_{ij,r} = \gamma_r$. When $k_{\text{tie}} > 0$, we set

$$\eta_{ij,r} = (H_r)_{ij}, \qquad H_r = G_r \Phi^{\top} + \Phi G_r^{\top}, \qquad (\Phi)_{a\ell} = \sqrt{\frac{2}{K}} \cos\left[\frac{\pi(2a-1)(2\ell-1)}{4K}\right],$$

where $G_r, \Phi \in \mathbb{R}^{K \times k_{\text{tie}}}$ and Φ contains the first k_{tie} columns of the fixed DCT-IV basis. This parameterization makes H_r symmetric while allowing tie propensity to vary across pairs. We do not use the covariance extension.

Rank selection and refitting. We initially assign each observation independently to training with probability 0.9, moving one validation observation from its most frequent outcome category to training whenever a model pair has no training observations. For the r-th objective in the configured order, we use seed $s_{\mathcal{D}} + r - 1$, with base seeds $s_{\mathcal{D}} = 2026071501, 2026071502, 2026071503$ for Arena-35, SearchArena-8, and VisionArena-12, respectively.

We evaluate

$$k_{\text{tie}} \in \{0, 1, 2, 3, 5\} \quad \text{for Arena-35,}$$

and

$$k_{\text{tie}} \in \{0, 1, 2\} \quad \text{for SearchArena-8 and VisionArena-12.}$$

The broader grid is used for the larger Arena-35 model set, while the factor rank is kept smaller for the two smaller model sets to avoid approaching a near-pairwise parameterization of the tie propensities.

For each candidate k_{tie}, we fit the training counts by multinomial maximum likelihood using BFGS with analytic gradients, with an L-BFGS-B fallback only after numerical failure. We select the successful candidate with the smallest held-out mean negative log likelihood, breaking an exact tie in favor of the smaller factor rank, and then refit the selected rank to all observations. All reported full-data fits converged with BFGS. The candidate and selected ranks, listed in configured-objective order, are reported in Tab. 13.

Table 13: Factored-Davidson tie-rank selection on the fixed seeded split described above. Selected ranks are listed in the configured-objective order given in App. D.1.

Dataset	Candidate k_{tie}	Selected k_{tie} by objective
Arena-35	$\{0, 1, 2, 3, 5\}$	$(2, 1, 0, 0, 0)$
SearchArena-8	$\{0, 1, 2\}$	$(2, 0, 2, 0, 0, 1, 2, 1, 0)$
VisionArena-12	$\{0, 1, 2\}$	$(2, 2, 1, 2, 1, 0, 1, 2)$

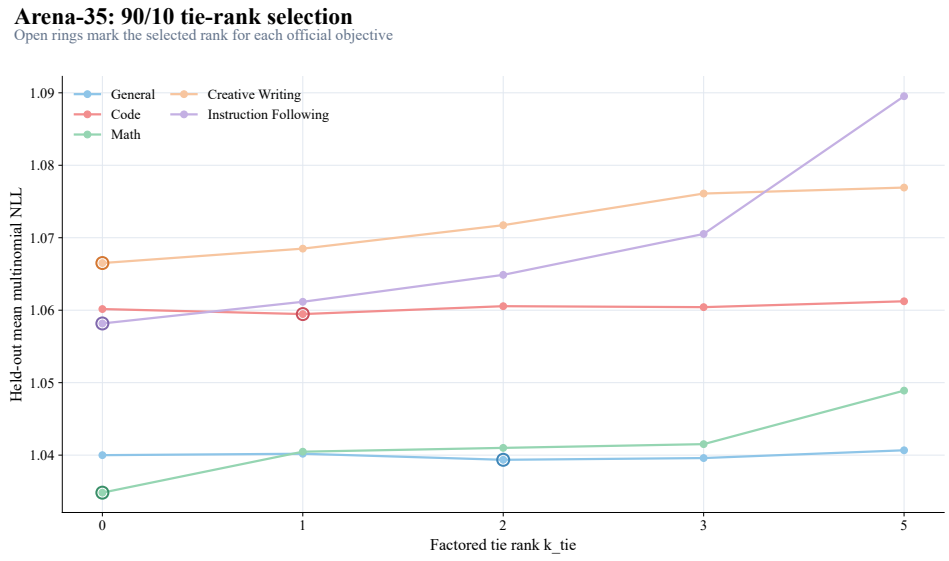


Figure 17: Arena-35: held-out NLL over candidate tie ranks; open circles mark the selected rank for each objective.

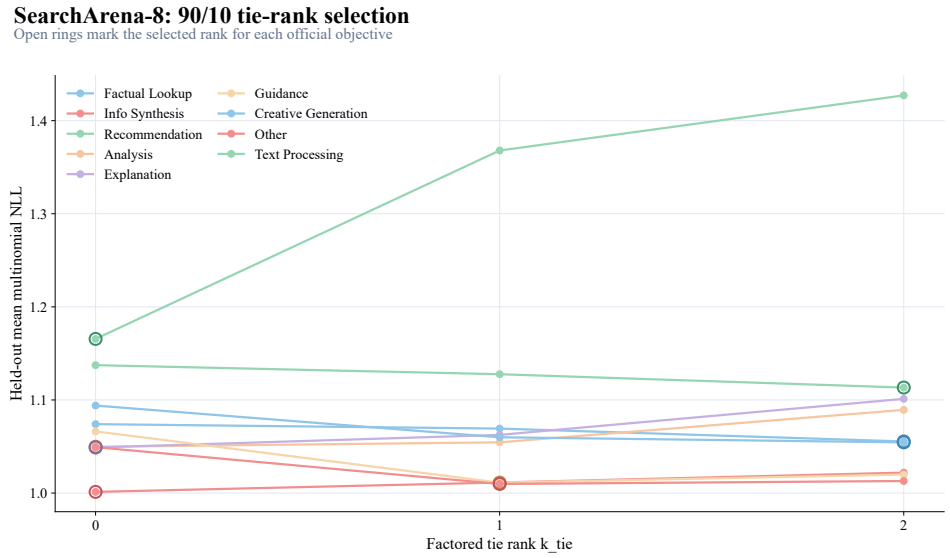


Figure 18: SearchArena-8: held-out NLL over candidate tie ranks.

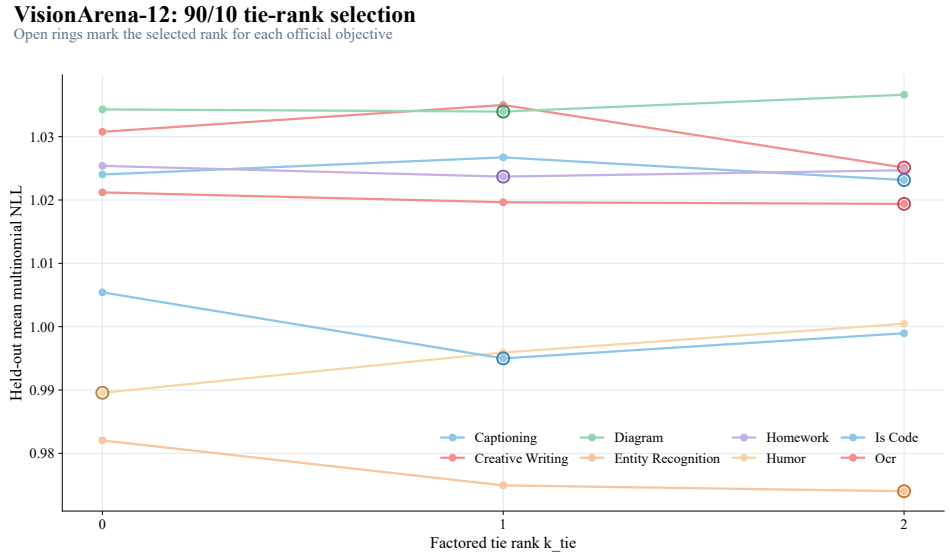


Figure 19: VisionArena-12: held-out NLL over candidate tie ranks.

The selected ranks vary across objectives and datasets in this fixed split, indicating that a single global tie rank need not be adequate. Because rank selection uses one seeded split, these results are treated as a reproducible model-selection step rather than an assessment of rank-selection uncertainty.

D.4.3 Factored-Davidson fitted target and comparison

Binary interface and fitted target. After the selected model is refitted on all observations, we convert its three fitted outcome probabilities to the complementary half-credit mean

$$q_{ij,r} = \pi^{W}_{ij,r} + \frac{1}{2}\pi^{M}_{ij,r}, \qquad q_{ji,r} = \pi^{L}_{ij,r} + \frac{1}{2}\pi^{M}_{ij,r} = 1 - q_{ij,r}.$$

Equivalently, an oracle query first draws from the fitted three-category distribution and resolves a fitted merged tie with an independent fair coin; it may therefore be implemented directly as a

Bernoulli draw with mean $q_{ij,r}$. We then compute

$$b^F_{i,r} = \frac{1}{K-1} \sum_{j \neq i} q_{ij,r}$$

and obtain the Pareto set of the fitted Borda scores $\mathcal{P}_F$. This is a fitted uniform-opponent Borda target, not an assertion that the Pareto set of the latent Davidson scores is the target.

VB-EGE **rerun.** For every dataset, we run 100 independent VB-EGE runs against one fixed comparison model fitted using Davidson, with $\delta = 0.05$ and $C = (2, 4, 4)$. Relative to the main experiment using only comparisons with a winner D, only the pair–objective oracle means, and hence the target coordinate matrix, change. The phase schedule, confidence radii, empirical-gap calculation, non-Pareto-first tie-breaking rule, and acceptance and elimination operations are unchanged.

These repetitions measure algorithmic variation conditional on the fitted comparison model F. They are not data bootstraps and do not propagate first-stage Davidson estimation or rank-selection uncertainty. The direct half-credit construction H remains a nonparametric target diagnostic and is not supplied to a separate VB-EGE rerun. Therefore, stopping-cost comparisons in this section apply only to D and F, whereas Pareto-target comparisons are reported for both D versus H and D versus F.

Table 14: Target sensitivity on the fixed sets of retained models used in the main experiments. $D/H/F$ give the strict-frontier sizes under decisive-only, direct merged half-credit, and fitted Davidson half-credit targets. $J(D, H)$ and $J(D, F)$ are Pareto-set Jaccard overlaps. Mean coordinate shifts are averaged over all model–objective coordinates, and $\rho_{\min}$ is the minimum objective-wise Spearman correlation with the decisive-only target.

Dataset	$D/H/F$	$J(D,H)$	$J(D,F)$	Mean $\lvert b^H - b^D \rvert$	Mean $\lvert b^F - b^D \rvert$	$\rho_{\min}(D,H)$	$\rho_{\min}(D,F)$
Arena-35	3/4/3	0.750	0.500	0.0244	0.0267	0.993	0.973
SearchArena-8	5/5/5	1.000	1.000	0.0333	0.0352	0.929	0.810
VisionArena-12	3/3/3	1.000	1.000	0.0529	0.0524	0.958	0.951

Under direct half-credit recoding H, Arena-35 adds `minimax-m1` to its decisive-only frontier and removes no model. Under the fitted Davidson target F, Arena-35 adds `minimax-m1` and removes `chatgpt-4o-latest-20250326`. SearchArena-8 and VisionArena-12 retain their decisive-only frontier membership under both alternatives. Thus, equal frontier sizes do not imply equal membership, and unchanged membership does not imply unchanged model–objective coordinates.

Table 15: VB-EGE behavior under sampling from logged decisive comparisons D and the fitted Davidson comparison model F. For $X \in \{D, F\}$, $\bar{\tau}_X$ is the sample mean stopping cost across 100 runs, reported in millions of comparisons. The cost ratio is the ratio of these sample means. The direct half-credit target H is not included because VB-EGE was not rerun against that target. All D and F runs stopped and exactly recovered their respective strict frontiers in $100/100$ repetitions.

Dataset	$\Delta^D_{\min}$	$\Delta^F_{\min}$	$\bar{\tau}_D$ (M)	$\bar{\tau}_F$ (M)	$\bar{\tau}_F/\bar{\tau}_D$	Exact recovery D/F
Arena-35	0.00441	0.00391	86.82	196.97	2.27×	100/100; 100/100
SearchArena-8	0.00460	0.00506	124.34	126.50	1.02×	100/100; 100/100
VisionArena-12	0.02171	0.00225	8.80	471.55	53.59×	100/100; 100/100

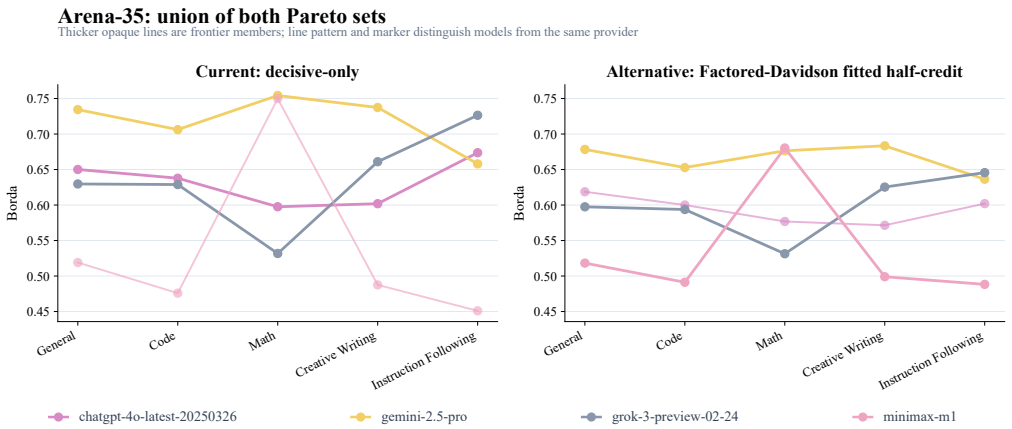


Figure 20: Arena-35: profiles of the union of the decisive-only and fitted Davidson Pareto sets, evaluated under each of the two targets.

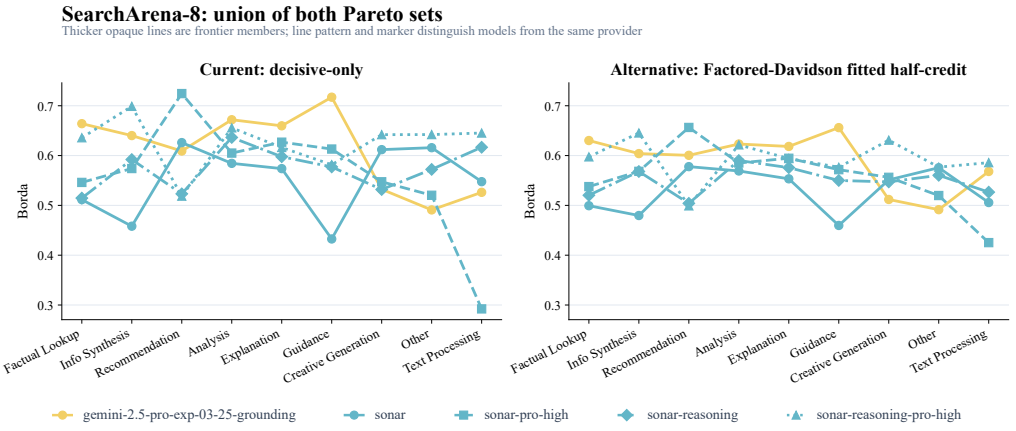


Figure 21: SearchArena-8: profiles of the union of the decisive-only and fitted Davidson Pareto sets, evaluated under each of the two targets.

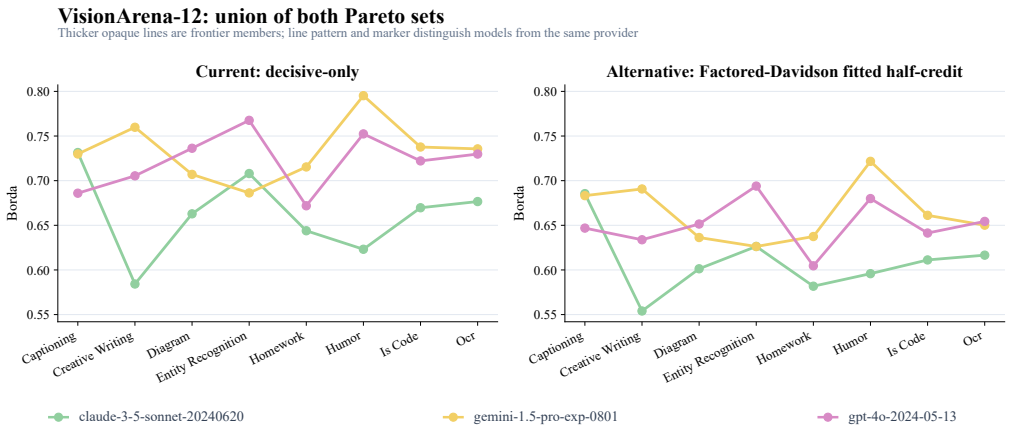


Figure 22: VisionArena-12: profiles of the union of the decisive-only and fitted Davidson Pareto sets, evaluated under each of the two targets.

E ADDITIONAL COMPARISON WITH TRACK-AND-STOP EXTENSIONS

E.1 IMPLEMENTATION USED IN THE REPORTED EXPERIMENTS

The original BT-MLE benchmark allocates uniformly over pair–objective tuples. The scalar method of Goldberger & Rudi (2026) instead combines Bradley–Terry MLE, adaptive allocation, and a generalized likelihood-ratio (GLR) stopping rule. We retain that template but replace its scalar top-k alternative by the following Pareto heuristic; this extension is ours, and the scalar theorem does not cover it. All reported experimental settings have $K \geq 16$ and therefore use the approximation below: first select a set of constraints using approximate Fisher-information scores, then fit the BT model subject to those constraints. We refer to this selection step as screening. The code invokes exact certificate enumeration only when $K \leq 8$ and $d^{|\widehat{\mathcal{P}}_t|} \leq 4096$.

At stopping check m, let t be the cumulative number of comparisons and $N_{\mathrm{cell}} = d\binom{K}{2}$ the number of pair–objective tuples. Write $\widehat{\theta}_{\cdot,r,t} = (\widehat{\theta}_{i,r,t})_{i=1}^{K} \in \mathbb{R}^K$ for the objective-r score vector and $\widehat{\theta}_t = (\widehat{\theta}_{i,t})_{i=1}^{K}$, where $\widehat{\theta}_{i,t} = (\widehat{\theta}_{i,r,t})_{r=1}^{d}$, for the collection of model score vectors. We use the same indexing for the constrained estimate $\widetilde{\theta}_t$. The implementation performs:

1. **Initialize and fit.** It first samples every pair–objective tuple twice. For each objective r, it then retains all opponent identities and fits

$$\widehat{\theta}_{\cdot,r,t} \in \arg \max_{\vartheta \in \mathbb{R}^K : \mathbf{1}^\top \vartheta = 0,\ \|\vartheta\|_\infty \leq 2} \ell_{r,t}(\vartheta),$$

   where $\ell_{r,t}$ is the Bradley–Terry log likelihood for objective r based on comparisons collected through query t. Fits use tolerance 10^{-7} and at most $1{,}000$ iterations.

2. **Construct constraints that change the Pareto set.** We call each such set of constraints a *certificate*. A drop certificate requires another current Pareto model to score at least as high on every objective; an add certificate makes an excluded model non-dominated by the current Pareto models. Let $\widehat{\mathcal{P}}_t = \mathcal{P}(\widehat{\theta}_t)$. A drop certificate (p, p'), for distinct $p, p' \in \widehat{\mathcal{P}}_t$, imposes $\lambda_{p',r} \geq \lambda_{p,r}$ for every r. An add certificate for $q \notin \widehat{\mathcal{P}}_t$ is an assignment $a : \widehat{\mathcal{P}}_t \to [d]$ imposing $\lambda_{q,a(p)} \geq \lambda_{p,a(p)}$ for every $p \in \widehat{\mathcal{P}}_t$. Thus one add arm has $d^{|\widehat{\mathcal{P}}_t|}$ possible assignments.

3. **Select constraints and refit the model.** Let $N_{ij,r}(t)$ count comparisons of the unordered pair $\{i, j\}$ on objective r, and write $\widehat{p}_{ij,r,t} = \sigma(\widehat{\theta}_{i,r,t} - \widehat{\theta}_{j,r,t})$. The observed Fisher information matrix is

$$H_{r,t} = \sum_{i<j} N_{ij,r}(t)\widehat{p}_{ij,r,t}(1 - \widehat{p}_{ij,r,t})(e_i - e_j)(e_i - e_j)^\top.$$

   Define

$$h_{ab,r,t} = \frac{\left((\widehat{\theta}_{b,r,t} - \widehat{\theta}_{a,r,t})^+\right)^2}{2(e_a - e_b)^\top H_{r,t}^\dagger (e_a - e_b)}.$$

   Here $e_a \in \mathbb{R}^K$ is the ath standard basis vector and $H_{r,t}^\dagger$ is the Moore–Penrose pseudoinverse of $H_{r,t}$. The screen scores a drop (p, p') by $\sum_r h_{p'p,r,t}$, and an add arm q by $s_q = \max_{p \in \widehat{\mathcal{P}}_t} \min_r h_{qp,r,t}$. For each add arm, the code pools the coordinate-wise cheapest assignment, all d constant-coordinate assignments, and assignments retained from the previous check; it ranks them by $\sum_p h_{qp,a(p),t}$, retains eight, and uses the cheapest. It then selects the globally smallest screened drop or add candidate. Only for this selected certificate $\widehat{a}_t$, it jointly solves the original constrained BT likelihood in each objective and sums numerical lower bounds on the log-likelihood loss imposed by these constraints to obtain G_t^{cert}. The implemented statistic is

$$Z_t = \min\{S_t, G_t^{\mathrm{cert}}\},$$

   where S_t is the selected global screening score.

4. **Allocate the next batch.** Let $\widetilde{\theta}_t$ be the constrained MLE for $\widehat{a}_t$. For each pair–objective tuple, compute

$$I_{ij,r,t} = \mathrm{kl}\Big(\sigma(\widehat{\theta}_{i,r,t} - \widehat{\theta}_{j,r,t}), \sigma(\widetilde{\theta}_{i,r,t} - \widetilde{\theta}_{j,r,t})\Big).$$

Define

$$I_{\max,t} := \max\left\{10^{-15}, \max_{\substack{1\le a<b\le K\\ s\in[d]}} I_{ab,s,t}\right\}.$$

Starting from uniform weights, we update

$$w_{ij,r}^{\text{new}} = \frac{w_{ij,r}^{\text{old}} \exp(0.8 I_{ij,r,t}/(\sqrt{m}\, I_{\max,t}))}{\sum_{a<b}\sum_{s=1}^{d} w_{ab,s}^{\text{old}} \exp(0.8 I_{ab,s,t}/(\sqrt{m}\, I_{\max,t}))}.$$

The next check targets

$$t_{\text{next}} := \min\{10^{18}, \max(t + N_{\text{cell}}, \lceil 1.8t\rceil)\}.$$

The normalized weights are mixed with uniform mass $\rho_t = t_{\text{next}}^{-1/3}$. C-tracking (Garivier & Kaufmann, 2016) allocates queries to pair–objective tuples whose observed counts fall short of their cumulative target counts, with at most 96 checks.

5. **Stop.** Define the comparison-count matrix

$$L_{r,t} = \sum_{i<j} N_{ij,r}(t)(e_i - e_j)(e_i - e_j)^\top.$$

The stopping threshold is

$$\beta_t(\delta) = \log\frac{1}{\delta} + \frac{0.1}{2}\sum_r \|\widehat{\theta}_{\cdot,r,t}\|_2^2 + \frac{1}{2}\sum_r \log\det(I_K + 2.5 L_{r,t}),$$

where I_K is the $K \times K$ identity matrix. The run stops only if the unconstrained fits and the fits under the selected constraints converge and $Z_t \ge \beta_t(\delta)$, and returns $\widehat{\mathcal{P}}_t$; otherwise it samples the tracked batch and repeats.

E.2 Comparison with the exact stopping statistic

Let $\mathcal{A}_t$ be all legal frontier-changing certificates. For $a \in \mathcal{A}_t$, let $\Lambda_{a,t}$ be the exact log-likelihood loss from imposing certificate a, summed over objectives; both the unconstrained and constrained optima use the centering and box constraints above. Let $\underline{\Lambda}_{a,t} \le \Lambda_{a,t}$ be the numerical lower bound returned by the solver. The exact stopping statistic is

$$Z_{\text{full},t} = \min_{a\in\mathcal{A}_t} \Lambda_{a,t}.$$

The implementation uses $G_t^{\text{cert}} = \underline{\Lambda}_{\widehat{a}_t,t}$, so its statistic is

$$Z_t = \min\{S_t, \underline{\Lambda}_{\widehat{a}_t,t}\}.$$

Screening based on Fisher information need not select the constraints with the smallest exact log-likelihood loss, so Z_t and $Z_{\text{full},t}$ have no deterministic ordering. An upward error can stop early; a downward error can delay stopping. A different selected certificate also changes the allocation and all later MLEs, frontiers, and stopping times. Equality at a check is guaranteed if the selected certificate gives the smallest exact log-likelihood loss, its numerical bound is tight, and $S_t \ge Z_{\text{full},t}$; screening does not guarantee these conditions. Hence all reported runs are labeled non-certified `screened-joint-certificate`; zero observed errors among stopped runs establish neither equivalence nor δ-correctness.

When $|\widehat{\mathcal{P}}_t| = 1$, full enumeration requires only

$$N_{\text{cert}} = (K-1)d,$$

at most 630 sets of constraints on the scaling grid. The scaling experiments should therefore be rerun on the same instances and random seeds using all constraint sets, or the stopping decision should be checked against all constraint sets with a theoretically valid stopping rule. For larger Arena and Witness fronts, add-certificate assignments alone can grow as $d^{|\widehat{\mathcal{P}}_t|}$, with constrained likelihood optimization required for each assignment, making naive global enumeration prohibitive. A practical Pareto MLE method could use Fisher scores to choose comparisons while relying on a theoretically valid stopping rule that accounts for all constraint sets.

E.3 PROVISIONAL SCREENED RESULTS

We report only the fixed-confidence comparisons at $\delta = 0.05$. Reusing the instances, seeds, and repetitions of App. C gives 12,400 runs per method over 28 experimental settings. Across scaling, benchmarks, and Pareto-set-size sensitivity, the screened heuristic uses fewer comparisons on average than VB-EGE in 1 of 12, 0 of 8, and 0 of 8 settings, respectively. The geometric means of its ratios of mean comparison counts to VB-EGE are 4.10, 251, and 20.6. For each setting s, we compute $R_s = \bar{\tau}_{\text{heuristic},s}/\bar{\tau}_{\text{VB-EGE},s}$. Across S settings, the reported geometric mean is $\exp(S^{-1}\sum_{s=1}^{S}\log R_s)$. For runs that do not stop before the experimental limit, the recorded comparison count is used; these summaries therefore describe the truncated runs, not their unknown stopping times.

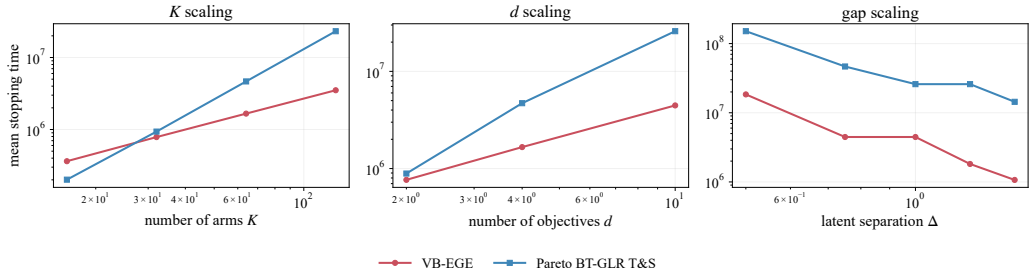


Figure 23: Provisional screened scaling at $\delta = 0.05$. The heuristic uses fewer comparisons on average than VB-EGE only at $K = 16$, and more in the other 11 settings. Curves show mean $\pm$ one SE.

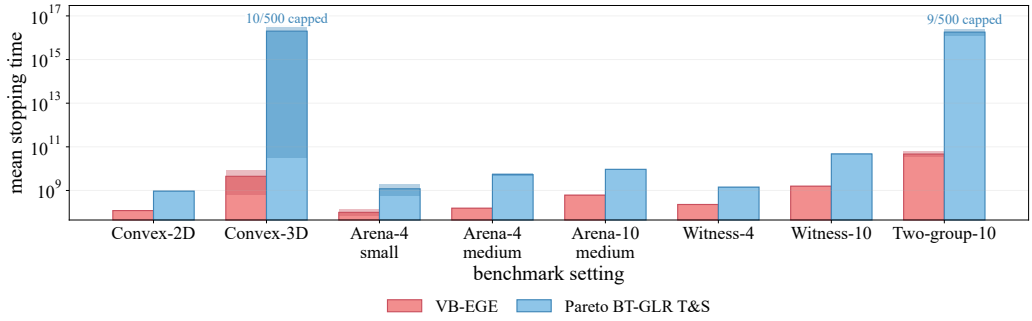


Figure 24: Provisional screened-implementation comparison at $\delta = 0.05$: eight benchmarks. Bars show means with one-standard-error summaries. The heuristic has 10 Convex-3D runs and 9 Two-group-10 runs that reach the experimental limit without stopping.

Overall, the screened heuristic uses fewer comparisons on average than VB-EGE in 1 of 28 settings, has a geometric mean of the ratios of mean comparison counts (heuristic/VB-EGE) of 21.1, and fails to stop in 19 benchmark runs, while all VB-EGE runs stop. These values do not characterize Track-and-Stop with exact optimization over all constraint sets: the scaling results remain provisional pending this check of the stopping decisions, and a computationally feasible adaptive MLE baseline with a proven fixed-confidence guarantee remains to be developed.

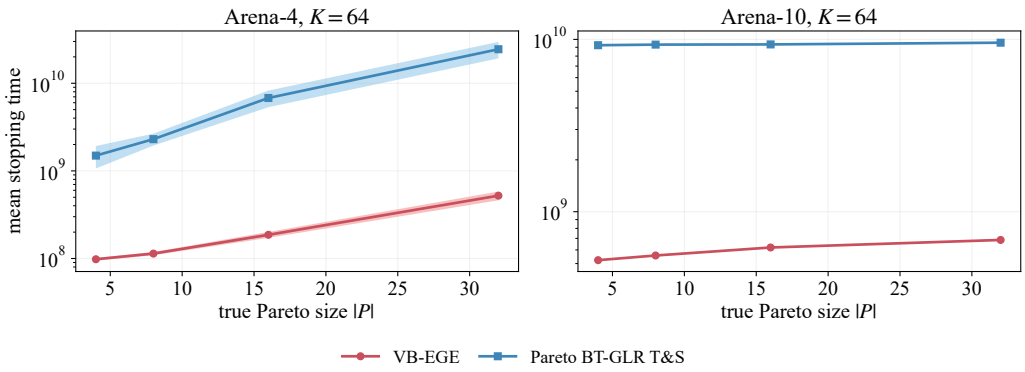


Figure 25: Pareto-set-size sensitivity at $\delta = 0.05$. The screened heuristic uses more comparisons on average than VB-EGE in all eight settings; the geometric mean of the ratios of mean comparison counts (heuristic/VB-EGE) is 20.6. Curves show mean $\pm$ one SE.