How Many Votes Make a Valid Leaderboard?

Anonymous authors
Paper under double-blind review

Abstract

Pairwise-vote leaderboards rank language models, but their displayed distinctions may exceed what the votes establish. We ask which leaderboard claims are statistically supported and which comparisons would resolve the remaining uncertainty. Our framework treats the best model, a top-k set, and a quality–cost frontier as separate claims. Under the Bradley–Terry model, each claim is tested against parameter regions supporting a different answer. We measure evidence by the smallest likelihood loss required to reach any such region. Tractable convex decompositions and conservative covers make these tests computationally feasible. We derive an anytime-valid stopping rule controlling false certification under adaptive sampling and stopping. To select new comparisons, we approximate the likelihood loss through the comparison graph's information geometry. Our algorithm, CERTIFY, targets comparisons that strengthen evidence against the hardest remaining alternatives. An audit of 1.16 million Chatbot Arena votes across 129 models reveals a gap between stability and certification. Split-half rankings remain highly correlated, yet none of the 63 adjacent top-64 positions clears our proved threshold. Even 355,507 comparisons among the top 20 provide only 84% of the evidence required to certify the leader. Thus, a reproducible ordering need not establish its individual positions. On held-out recorded votes for a 16-model subset, CERTIFY identifies the leader in a median 12,480 comparisons. Across 15 runs at a nominal 5% error level, every returned answer is correct. Under the same stopping rule, competing sampling strategies require approximately 5.6–7.5x as many votes. Experiments also examine quality–cost and two-attribute frontiers, extending the approach beyond identifying a single winner. Together, these results connect the statistical support for leaderboard claims to the comparisons needed to establish them.

1 Introduction

Pairwise-vote leaderboards compress a large, irregular comparison archive into an ordered list. That list is easy to read, but can imply distinctions that the underlying evidence does not resolve (Card et al., 2020; Zhou et al., 2026; Oyarhoseini et al., 2026). It asserts not only that the leading model has the largest fitted score but also that every displayed position is distinguishable from the next. These claims need not be equally well supported: a leaderboard can be stable in broad outline, while its adjacent ranks remain statistically unresolved (Zhou et al., 2026; Neuhof and Benjamini, 2026). This distinction matters for evaluations, such as Chatbot Arena (Chiang et al., 2024). In general, a growing literature documents the fragility and limited separability of benchmark rankings (Dehghani et al., 2021; Perlitz et al., 2024; Zhou et al., 2026; Oyarhoseini et al., 2026). Specifically, performance can shift with the selected tasks (Dehghani et al., 2021). A benchmark leader can change when an apparently irrelevant low-ranked model is removed (Perlitz et al., 2024). In addition, item-response analyses find poor separability among leading models even when aggregate ranks appear stable (Zhou et al., 2026). Small targeted perturbations can also change the winner and substantially alter rank correlation and reported confidence intervals (Oyarhoseini et al., 2026), and these concerns about low statistical power long predate LLM leaderboards (Card et al., 2020).

These findings mostly establish that rankings can be unstable but instability and evidential sufficiency are different questions. *Stability analysis is counterfactual*: it asks how much the data or evaluation design must change before the reported ranking changes (Dehghani et al., 2021; Oyarhoseini et al., 2026). *Statistical certification is inferential*: it asks whether the collected observations

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.

rule out every parameter region in which the reported answer is false, while controlling the probability of an incorrect conclusion (Garivier and Kaufmann, 2016; Degenne and Koolen, 2019). Mixture martingales and confidence sequences provide the corresponding error control under adaptive sampling and stopping (Kaufmann and Koolen, 2021; Howard et al., 2021). A position may therefore be insensitive to a prescribed class of perturbations and still lack enough evidence for certification. Conversely, a statistically certified claim need not be invariant to arbitrary changes in the benchmark. Our goal is the latter, i.e., *to attach a stated error level to the claims a leaderboard reports*.

Standard diagnostics do not directly provide such a certificate. Per-model confidence intervals quantify uncertainty in individual fitted scores rather than testing the composite event that a reported ranking is wrong. Tests that use only the direct votes between two models discard information transmitted through their comparisons with third models. Split-half rank correlation measures global reproducibility but it can remain high when individual positions are unresolved (). Statistical works on pairwise comparisons explains that estimation accuracy depends on the topology of the comparison graph (Shah et al., 2016), and active-ranking procedures must allocate samples to the comparisons relevant to the target ordering (Heckel et al., 2018). Recent methods improve LLM ranking through prediction-powered inference, or structural assumptions like sparse comparisons (Chatzi et al., 2024; Li et al., 2026). Our question is complementary.

Given the votes and the fitted leaderboard, which of its reported claims are established, and which comparison should be collected next when they are not?

We study this question under the Bradley–Terry model (BT) (Bradley and Terry, 1952), which is also used to aggregate pairwise preferences in Chatbot Arena (Chiang et al., 2024). A leaderboard answers claims, such as the identity of the top model, a top-k model set, or a quality–cost frontier. The answer depends on the fit of the BT model $\widehat{\theta}$ and the alternative parameters acting as witnesses against correctness of a claim. The evidence against a witness region A is the exact constrained likelihood drop $\Lambda_A(N) = \ell_N(\widehat{\theta}) - \sup_{\theta' \in A} \ell_N(\theta')$, where N is the number of comparisons. The region with the smallest drop identifies the most plausible evidence for the current claim to be wrong. This shifts the basic object of evaluation from uncertainty in an individual score to the likelihood cost of changing the reported answer. In structured identification problems, evaluating this drop requires projecting the fitted parameter onto competing witness regions with alternative answers. With multiple possible answers, the witness set is generally nonconvex and may require a separate constrained optimization for each competing answer (Degenne and Koolen, 2019; Li et al., 2024).

We observe that though witness set is generally nonconvex, the claims relevant to leaderboards admit compact decompositions into convex pieces. For top-1 (Garivier and Kaufmann, 2016; Degenne and Koolen, 2019) and top-k (Goldberger and Rudi, 2026), these pieces are half-spaces. For quality–cost and two-attribute frontiers (Kone et al., 2024; 2025; Shukla and Basu, 2024; Das et al., 2025b), they also include joint entry events in which a currently dominated model escapes all of its dominators. We derive tractable claim-specific decompositions and computationally cheaper convex covers. Because a cover may enlarge, but never omit, the witness set, it can delay certification but cannot create evidence that was not present. This ensures statistical validity of the final conclusions. We use the same witness geometry to direct new comparisons. For routing, we replace the exact likelihood drop by its Gaussian projection $S_A(N) = \frac{1}{2} \inf_{\theta' \in A} \|\theta' - \widehat{\theta}\|^2_{L(N)}$, where $L(N)$ is the information-weighted Laplacian of the comparison graph. The derivative of S_A with respect to an edge measures how much another vote on that pair is expected to strengthen the certificate. This results in an anytime valid sampling scheme for yielding statistically valid leaderboards.

The empirical consequences are sharp (Figure 1). We audit 1.16 million recorded Chatbot Arena votes. We find that while the broad ordering is reproducible, none of the 63 consecutive positions in the published top-64 ordering clears the proved threshold. The best-supported adjacent step reaches only 6.7% of the required evidence, and the median reaches 0.42%. Thus, a stable global ranking need not support fine-grained claims about individual positions. When new votes can be selected, the same analysis supplies a remedy: in held-out replay on recorded Arena votes, the proposed certificate-directed routing identifies the top model among the top 20 using 31% of the votes required by uniform routing under the same stopping rule. It succeeds by sampling comparisons relevant to the closest wrong answer, rather than pairs that are merely uncertain.

Our contributions are

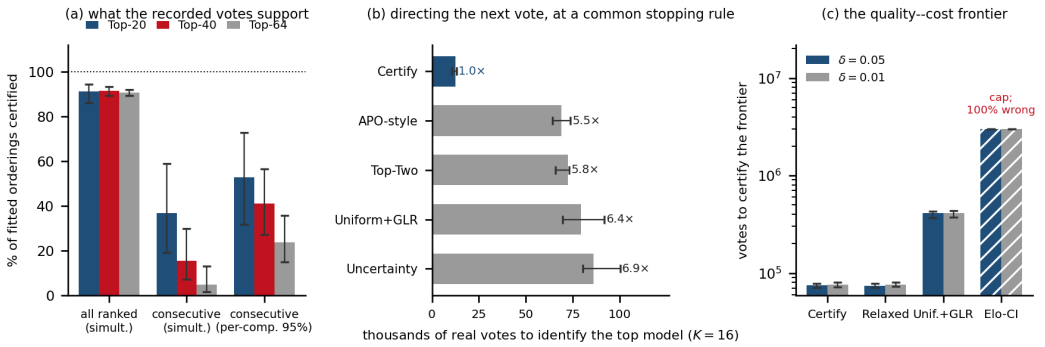

Figure 1: **What the votes support, what they cost, and what directing them buys.** *(a)* The fraction of fitted pairwise orderings that the recorded Arena votes certify, for the top 20, 40 and 64 models. *(b)* Votes needed to identify the top model, replaying held-out recorded votes (15 seeds, $\delta = 0.05$); Routing by the witness price is 2.8–5.8× cheaper than every alternative, and the ordering of the baselines is itself informative: uncertainty-directed sampling is the closest competitor and still 2.8× behind. *(c)* Votes needed to certify the simulated quality–cost frontier (25 seeds), on a log scale. Hatched bars did not finish: ELO-CI reached the three-million-vote cap in all 25 runs and named the wrong frontier in all 25, and both are about 5.5× cheaper than uniform collection.

1. **A claim-level formulation with a price and a certificate.** Each statement a leaderboard makes a position, a top-k set, or a Pareto frontier over several attributes is treated as a claim, and each claim corresponds to the witness regions that it must rule out. We give tractable witness covers for top-1, top-k and for frontier identification. So, each claim is certified by the regions its own geometry requires. Further, a stopping rule that is δ-correct under any sampling rule, any witness set, and any stopping time, with no union bound over claims, regions or times (Theorem 3.2). We further show upper bound on sample complexity of the proposed rule (Theorem 3.3).

2. **An audit of the largest public leaderboard against its own votes: The ordering is not supported, while the ordering is stable.** On 1.16M Arena votes over 129 models, not one of the 63 consecutive positions among the top 64 is supported under the threshold we can prove (3 are under the weaker boundary our experiments run on). Yet 92.3% of *all* ranked pairs in the top 64 are certifiable, and splitting each pair's votes into disjoint halves and refitting leaves Kendall's τ at 0.978, 0.971 and 0.967 for the top-20, 40 and 64, with only 1.1–1.6% of pairs transposing and those are precisely the adjacent pairs the ordering asserts. The leaderboard is globally reliable and locally uninformative, and the two statistics measure the two halves of that sentence. Unlike perturbation and item-response diagnostics in prior work, the verdict is per position computed from the evidence already collected, and carries a stated confidence.

3. **Which vote to collect next, across claims and on a live leaderboard task.** Routing by the gradient of the near-binding witness price identifies the top model from held-out recorded Arena votes in 13,394 votes, against 77,614 for top-two sampling, 72,614 for APO-style active preference learning, 43,634 for uniform collection and 37,854 for uncertainty-directed sampling, correct in 15/15 runs; uncertainty-directed is the strongest baseline and still 2.8× behind, because the comparisons that look most uncertain are between models nobody doubts. The frontier claim costs 74,460 votes against 410,980 for uniform, while an Elo-interval rule is wrong in 25/25 runs and active Pareto exploration does not finish in three million. On twelve real policies spanning a helpful–harmless tradeoff judged on Anthropic's HH-RLHF prompts, certifying the frontier takes 2.00M votes with 0/10 errors, where random collection is wrong in 10/10 and a confidence-interval rule stops in 767,492 but is wrong in 7/10.

2 PROBLEM FORMULATION

Sampling and Ranking Models. There are K models with unknown Bradley–Terry scores $\theta \in \mathbb{R}^K$ and unknown expected cost associated with each model c. We assume that the scores are recentered to ensure $\sum_i \theta_i = 0$. The arm set is the set of *pairs* $\mathcal{A} = \{(i,j) : i < j\}$. $|\mathcal{A}| = \binom{K}{2}$. Pulling arm (i,j) returns a Bernoulli outcome $Z \sim \mathrm{Ber}\big(\sigma(\theta_i - \theta_j)\big)$ independently across pulls, the likelihood

is given by: $\ell_N(\theta) = \sum_{i \neq j} W_{ij} \log \sigma(\theta_i - \theta_i)$, W_{ij} is the number of times i beats j and $\sigma(x) = (1 + e^{-x})^{-1}$. An algorithm is a triple: a *sampling rule* $(A_t)_{t \geq 1}$ with $A_t \in \mathcal{A}$ measurable with respect to $\mathcal{F}_{t-1} = \Sigma\big((A_s, Z_s)_{s \leq t-1}\big)$; a *stopping rule* $S(\mathcal{F}_{t-1}, \delta)$ yielding a stopping time τ; and a *decision rule* $\hat{a}$.

Claim map. A claim is a map. For example, $\mathrm{ans}_{\text{top-1}}(\theta) = \arg\max_i \theta_i$, while $\mathrm{ans}_{\text{front}}(\theta, c) = \{\, i : \nexists\, j$ with $\theta_j > \theta_i$ and $c_j \leq c_i \,\}$. The second one incorporates a known cost $c \in \mathbb{R}^K$ to construct a quality-cost frontier. We observe that both are determined by finitely many linear inequalities in θ.

Definition 2.1. *An algorithm is δ-correct if for every θ, $\mathbb{P}_\theta\big(\tau < \infty,\ \hat{a} \neq \mathrm{ans}(\theta, c)\big) \leq \delta$.*

Goal: The the objective is to design a δ-correct algorithm that minimises $\mathbb{E}_{\theta,c}[\tau]$.

Sample complexity. Write $N_t \in \mathbb{R}^{\mathcal{A}}$ for pull counts and $w \in \Delta_{\mathcal{A}}$ for an allocation. The Kullback–Leibler transportation cost of moving θ to θ' under allocation w is $d(w; \theta, \theta') = \sum_{(i,j) \in \mathcal{A}} w_{ij}\, \mathrm{kl}\big(\sigma(\theta_i - \theta_j),\, \sigma(\theta'_i - \theta'_j)\big)$, and the change-of-measure argument (Garivier and Kaufmann, 2016) gives, for any δ-correct algorithm, $\liminf_{\delta \to 0} \frac{\mathbb{E}_\theta[\tau]}{\ln(1/\delta)} \geq T^\star(\theta)$ such that $(T^\star(\theta))^{-1} = \max_{w \in \Delta_{\mathcal{A}}} \inf_{\theta' \in \mathrm{Alt}(\theta)} d(w; \theta, \theta')$.

The alternative hypothesis set. A claim is an answer map $\mathrm{ans} : \mathbb{R}^K \to \mathcal{C}$ into a finite set, determined by finitely many linear inequalities in θ. Given the current fit, the hypothesis to be refuted is that the truth supports a *different* answer: $\mathrm{Alt}(\theta) = \{\theta' \in \mathbb{R}^K : \mathrm{ans}(\theta') \neq \mathrm{ans}(\theta)\}$. This is composite, and in general neither convex nor connected. It is, however, a finite union of convex polyhedra, and it is that decomposition the procedure uses: $\mathrm{Alt}(\theta) = \bigcup_{A \in \mathcal{W}(\theta)} A$, A convex, polyhedral. Two instances. For the top-1 claim with $a = \arg\max_i \theta_i$, each piece is a single half-space, $A_b = \{\theta' : \theta'_b \geq \theta'_a\}$, one per challenger model $b \neq a$. For the quality–cost frontier with known costs c, the pieces are the *exits* $X_{i,j} = \{\theta' : \theta'_j > \theta'_i\}$ for i on the frontier and $c_j \leq c_i$, together with the *entries* $E_m = \bigcap_{j:\, c_j \leq c_m,\ j \neq m} \{\theta' : \theta'_m \geq \theta'_j\}$, ($m$ off the frontier), which are intersections rather than half-spaces. The decomposition must be *complete*: every θ' with a different answer must lie in some piece. The inner infimum decomposes over the convex pieces of Alt, so the program is a maximin of finitely many concave functions of w.

Definition 2.2 (Transportation cost)**.** *For a convex $A \subseteq \mathbb{R}^K$ and counts N, $S_A(\theta, N) = \inf_{\theta' \in A} \sum_{i<j} n_{ij}\, \mathrm{kl}\,\big(\sigma(\theta_i - \theta_j),\, \sigma(\theta'_i - \theta'_j)\big)$.*

This is the cheapest way, in accumulated Kullback–Leibler divergence weighted by how often each pair was actually voted on, to move the fitted parameter into the alternative. It is concave and positively homogeneous of degree one in N, and finite whenever A is nonempty.

Claims and witnesses. Scores $\theta \in \Theta = \mathbb{R}^K$ are identified up to shift. A *claim map* ans assigns to each parameter the assertion it supports, by a formula in strict comparisons of scores; where the formula does not produce the claimed assertion it produces something else. For top-1, $\mathrm{ans}(\theta) = a$ iff $\theta_a > \theta_j$ for all $j \neq a$.

Definition 2.3 (Witness, witness set)**.** *Fix θ and let* $\mathrm{Alt}(\theta) = \{\theta' : \mathrm{ans}(\theta') \neq \mathrm{ans}(\theta)\}$. *The set of parameters that* robustly *leave the answer unchanged is the interior $\mathcal{U}(\theta) = interior(\{)\theta' : \mathrm{ans}(\theta') = \mathrm{ans}(\theta)\}$. A* witness against $\mathrm{ans}(\theta)$ *is any $\theta' \notin \mathcal{U}(\theta)$, and the* witness set *is $\mathcal{W}(\theta) = \overline{\mathrm{Alt}(\theta)} = \Theta \setminus \mathrm{int}\{\theta' : \mathrm{ans}(\theta') = \mathrm{ans}(\theta)\}$,*

The second equality is the elementary fact that the complement of an interior is the closure of the complement. Taking the closure admits exactly the tie points on the edge of $\mathrm{Alt}(\theta)$ and nothing else: a tie that does not affect the answer (two models both far from the top, say) stays out. The payoff is that $\mathcal{W}(\theta)$ is closed, so the minima below are attained, and no proof needs a tie clause.

Definition 2.4 (Cover)**.** *A* cover *of $\mathcal{W}(\theta)$ is a finite collection $\mathcal{C}$ of closed convex sets with $\mathcal{W}(\theta) \subseteq \bigcup_{C \in \mathcal{C}} C$. Its price is $\Lambda_{\mathcal{C}}(N) = \min_{C \in \mathcal{C}} \Lambda_C(N)$.*

Here, unlike for a decomposition, the inclusion carries information: it may be strict, and a cover may contain parameters that are not witnesses.

Lemma 2.5 (Covers are safe and cost only speed)**.** *For any cover, $\Lambda_{\mathcal{C}}(N) \leq \Lambda_{\mathcal{W}}(N)$. Equality holds when the minimiser of the cheapest $C \in \mathcal{C}$ is itself a witness, specifically when every $C \subseteq \mathcal{W}$.*

Algorithm 1 CERTIFY: certificate-directed vote routing

Require: answer map $\mathrm{ans}(\cdot)$; risk δ; ridge $\lambda > 0$; vote oracle; batch b; window exponent $\rho = 0.9$; servable pairs $\mathcal{P}_{\mathrm{live}}$
Ensure: a claim certified at level δ, or EXHAUSTED

init: cast 2 votes on each $(i, j) \in \mathcal{P}_{\mathrm{live}}$; $t \leftarrow$ total votes; $\hat{\theta} \leftarrow$ BT-MLE
loop
  $L \leftarrow L(N_t; \hat{\theta})$; $\mathcal{C} \leftarrow \textsc{Cover}(\mathrm{ans}, \hat{\theta})$ ▷ covers $\mathcal{W}(\hat{\theta})$; Algo. 2
  $\beta \leftarrow \ln \frac{1}{\delta} + \frac{1}{2} \ln \det_E\big(I + \frac{V_t}{4\lambda}\big) + \frac{\lambda}{2}\|\hat{\theta}\|^2$ ▷ one boundary, every piece
  for all $P \in \mathcal{C}$ **do**
    $\Lambda_P \leftarrow \textsc{ExactPrice}(P, \hat{\theta}, N_t)$ ▷ log-GLR; decides stopping
  if $\Lambda_P \geq \beta$ for every $P \in \mathcal{C}$ **then return** $\mathrm{ans}(\hat{\theta})$ ▷ Theorem 3.2

  for all $P \in \mathcal{C}$ **do**
    $(S_P, z_P) \leftarrow \textsc{QuadPrice}(P, L, \hat{\theta})$ ▷ Alg. 3; routing only
  $S^\star \leftarrow \min_{P \in \mathcal{C}} S_P$; $\mathcal{B} \leftarrow \{P \in \mathcal{C} : \ S_P \leq (1 + t^{-\rho}) S^\star\}$ ▷ near-binding pieces
  $G_{ij} \leftarrow \dfrac{1}{|\mathcal{B}|} \displaystyle\sum_{P \in \mathcal{B}} \tfrac{1}{2}\, \sigma'(\hat{\theta}_i - \hat{\theta}_j)\, \big((z_P)_i - (z_P)_j\big)^2$ ▷ Danskin: no re-solve

  if $\min_{(i,j) \in \mathcal{P}_{\mathrm{live}}} n_{ij} < \sqrt{t} \,/\, |\mathcal{P}_{\mathrm{live}}|$ **then**
    $(i^\star, j^\star) \leftarrow \arg\min_{(i,j) \in \mathcal{P}_{\mathrm{live}}} n_{ij}$ ▷ forced exploration
  else
    $(i^\star, j^\star) \leftarrow \arg\max_{(i,j) \in \mathcal{P}_{\mathrm{live}}} G_{ij}$ ▷ Frank–Wolfe vertex
  cast b votes on $(i^\star, j^\star)$; update N_t, V_t and t
  if $(i^\star, j^\star)$ is exhausted **then** $\mathcal{P}_{\mathrm{live}} \leftarrow \mathcal{P}_{\mathrm{live}} \setminus \{(i^\star, j^\star)\}$
  if $\mathcal{P}_{\mathrm{live}} = \emptyset$ **then return** EXHAUSTED
  $\hat{\theta} \leftarrow$ BT-MLE at N_t, warm-started

So a stopping rule that fires when $\Lambda_{\mathcal{C}}(N)$ clears the threshold fires only when $\Lambda_{\mathcal{W}}(N)$ does: pricing a cover can delay certification but can never make it wrong. The reverse is not true. A collection whose union *misses* witnesses is not a cover, can have a price above $\Lambda_{\mathcal{W}}$, and can certify while a cheaper witness was never examined. We make the following standard assumptions for the proposed method.

Assumption 2.6 (Interior truth)**.** $\theta \in \operatorname{int} \Theta_R$.

Assumption 2.7 (Stable claim)**.** *θ lies in the interior of $\{\vartheta : \mathrm{ans}(\vartheta) = \mathrm{ans}(\theta)\}$. Equivalently, all the gaps that define the claim are strict.*

Assumptions 2.6 is used for the upper bound of Theorem 3.3. Assumption 2.7 gives the convergence arguments a fixed objective. Theorem 3.2 uses none of the four: it needs only completeness.

3 CERTIFICATE-DIRECTED VOTE ROUTING

Let $N_t = (n_{ij}(t))$ be the comparison counts after t votes and $\hat{\theta}_t$ the Bradley–Terry maximum likelihood estimate, under which $\mathbb{P}_{\hat{\theta}_t}(i \succ j) = \sigma(\hat{\theta}_{t,i} - \hat{\theta}_{t,j})$. The candidate answer is $\hat{a}_t = \mathrm{ans}(\hat{\theta}_t)$. Algorithm 1 repeats three steps: fit the scores, test whether $\hat{a}_t$ is already certified, and if not direct the next batch of votes at the comparisons that most reduce the remaining doubt. It does not estimate every score equally well; it concentrates on the comparisons that separate the current claim from its most plausible alternatives. It begins by casting two votes on every servable pair, which connects the comparison graph and steadies the first fit.

Representing the ways the claim can fail. $\textsc{Cover}(\mathrm{ans}, \hat{\theta}_t)$ returns a finite family $\mathcal{C}_t$ of closed convex sets with $\mathcal{W}(\hat{\theta}_t) \subseteq \bigcup_{P \in \mathcal{C}_t} P$, where $\mathcal{W}(\hat{\theta}_t)$ is the set of parameters under which the current answer changes. Each piece is one way $\hat{a}_t$ could be false. For a top-1 claim with winner a there is one half-space $\{\theta' : \theta'_j \geq \theta'_a\}$ per challenger; for a top-k set S, one per pair $a \in S$, $j \notin S$. These covers are exact, with $K - 1$ and $k(K - k)$ pieces. For frontier claims the answer can change either by a member exiting or by a dominated model entering, so the cover carries *exit* pieces for current

members and *entry* pieces for currently dominated models, retaining only current dominators. In the quality–cost case exits are half-spaces and each entry is an intersection of inequalities against the entrant's dominators; in the two-attribute case an exit needs a competitor to dominate on both axes, and entry pieces enumerate which axis is used to escape each dominator.

Exact certification. For each $P \in \mathcal{C}_t$, EXACTPRICE computes the log generalised likelihood ratio $\Lambda_P(N_t) = \min_{\theta' \in P} \left[\ell_{N_t}(\hat{\theta}_t) - \ell_{N_t}(\theta')\right]$, the smallest log-likelihood loss incurred by moving from the fit into P. A large value means even the best-fitting parameter in P explains the observed votes substantially worse than $\hat{\theta}_t$. The constrained minimisation enumerates candidate active sets, optimises the exact negative log-likelihood under the corresponding equality constraints, and keeps candidates satisfying the primal and dual KKT conditions, convexity of the negative log-likelihood, and linearity of the constraints make these checks identify the constrained minimum. For a top-1 or top-k half-space, this is a single Newton step to solve on the boundary $\theta'_j = \theta'_a$.

Definition 3.1 (Stopping rule)**.** *Fix $\delta \in (0,1)$ and $\lambda > 0$ in advance, let $V_t = \sum_{s \le t} x_s x_s^\top$ be the design Gram matrix the unweighted Laplacian of the vote counts and set $\beta(t,\delta) = \ln \frac{1}{\delta} + \frac{1}{2} \ln \det_E \left(I + \frac{V_t}{4\lambda}\right) + \frac{\lambda}{2} \|\hat{\theta}_t\|^2$, the determinant taken on $\mathbf{1}^\perp$. Stop at $\tau = \inf \left\{t \ge 1 : \hat{\theta}_t \text{ exists and } \Lambda_P(N_t) \ge \beta(t,\delta) \text{ for every } P \in \mathcal{C}_t\right\}$, and return $\mathrm{ans}(\hat{\theta}_\tau)$.*

Certification is the statement that *every* way the answer could be wrong has been made expensive; the binding piece is what the rule waits on. Three features are worth naming, because each is a place where a weaker rule would be tempting. *The same β serves every piece.* There is no $\delta/|\mathcal{C}_t|$ and no factor for the data-dependent active set: the proof bounds Λ_P at the true parameter by $\ell_t(\hat{\theta}_t) - \ell_t(\theta)$, an event that mentions no region at all, so one boundary covers the whole cover simultaneously. β *is random*, depending on the data through V_t and $\hat{\theta}_t$; the argument is pointwise in ω and never needs it deterministic.

Theorem 3.2 (δ-correctness)**.** *Under the Bradley–Terry model, for the rule of Def. 3.1, any complete witness system, and* any *sampling rule (adaptive or not): $\mathbb{P}_\theta(\tau < \infty) = 1$, and $\mathbb{P}_\theta\big(\mathrm{ans}(\hat{\theta}_\tau) \ne \mathrm{ans}(\theta)\big) \le \delta$.*

The proof is in Appendix F: a likelihood-ratio martingale, a Gaussian mixture over the $K-1$ free directions, and Ville's inequality. Completeness of the witness system is the only property of it used. The guarantee is simultaneous over times, regions and answer maps, the same votes may be used to test many claims, and with probability at least $1-\delta$ no false claim passes at any time, with no union bound over claims or stopping times (Remark F.7). Nothing in the choice of which vote to buy can compromise validity; routing affects only speed.

Identifying the certificate bottlenecks. When the stopping condition fails, the algorithm asks which alternatives are hardest to rule out. It forms the information operator $L_t = L(N_t; \hat{\theta}_t)$ and, for every $P \in \mathcal{C}_t$, calls QUADPRICE (Algorithm 3) for the information-metric projection $S_P = \frac{1}{2} \min_{\theta' \in P} \|\theta' - \hat{\theta}_t\|^2_{L_t}$ and the displacement z_P attaining it, by enumerating nonempty active subsets of the inequalities defining P and keeping primal- and dual-feasible candidates. For top-1 and top-k each piece has a single inequality and the enumeration has one case, giving the familiar $S_P = \mathrm{gap}^2/(2R_{ij})$ with R_{ij} the effective resistance; multi-constraint frontier pieces may need several candidates. For a two-attribute frontier the information matrix is block diagonal across the independently observed axes, so each projection is the sum of two axis-specific projections. S_P is used for routing only: it is a quadratic surrogate for Λ_P, related to it by the sandwich of Theorem F.10 and by nothing else.

Routing on the single smallest price alone is unstable when several pieces are nearly tied, so the algorithm sets $S^\star_t = \min_{P \in \mathcal{C}_t} S_P, \quad \mathcal{B}_t = \{P \in \mathcal{C}_t : S_P \le (1 + t^{-\rho}) S^\star_t\}, \quad \rho = 0.9.$ $\mathcal{B}_t$ holds the *near-binding* pieces. Early on the window is broad, which makes the allocation less sensitive to noise and near ties; as t grows it contracts onto the genuinely binding pieces.

Choosing the next comparison. For each live pair the routing score is $G_{ij}(t) = \frac{1}{|\mathcal{B}_t|} \sum_{P \in \mathcal{B}_t} \frac{1}{2} \sigma'(\hat{\theta}_{t,i} - \hat{\theta}_{t,j}) \left((z_P)_i - (z_P)_j\right)^2$. Its two factors read simply. The curvature $\sigma'(\hat{\theta}_{t,i} -$

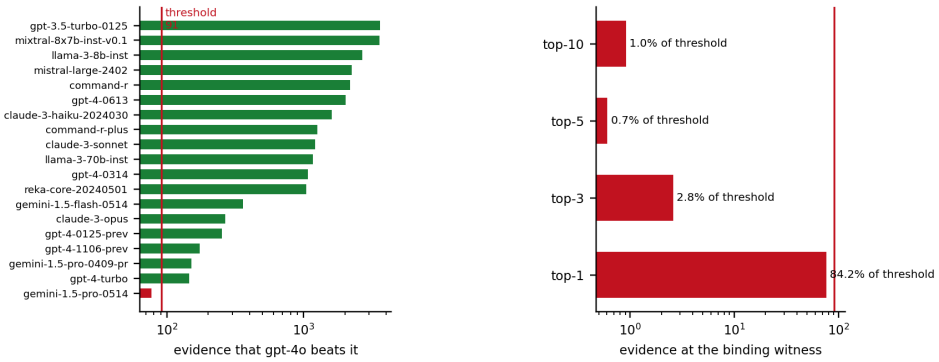

Figure 2: **Naming the best model is nearly within reach; naming the best three is not.** Both panels check claims about the top-20 Arena models against the 355,507 votes recorded among them. Left: for the claim that `gpt-4o-2024-05-13` is best, Sixteen clear the threshold comfortably; the claim is decided by the one that does not, `gemini-1.5-pro-0514` which is 84% of what is required. Right: the binding witness for the top-k claim, $k \in \{1, 3, 5, 10\}$. Top-1 reaches 84%; top-3 reaches 2.9%, top-5 0.7% and top-10 1.0%. Adding models to a leaderboard does not make its top-k claims easier; it adds cheaper ways for them to fail.

$\hat{\theta}_{t,j}$) is the information a single Bradley–Terry comparison supplies, largest when the two are close in estimated strength. The squared contrast $((z_P)_i - (z_P)_j)^2$ measures how relevant the pair is to the adverse direction associated with P. A pair therefore scores highly when it is both informative and pointed at the pieces preventing termination. The expression follows from Danskin's theorem: once z_P is in hand the directional derivative of the optimal quadratic price may be evaluated at z_P itself, so all pairs are scored without re-solving QUADPRICE. Absent an exploration step the algorithm takes $(i_t^\star, j_t^\star) \in \arg\max_{(i,j)\in\mathcal{P}_{\text{live}}} G_{ij}(t)$.

Forced exploration and exhausted pairs. A purely greedy rule could abandon a pair on the strength of a poor early estimate. The algorithm therefore checks whether $\min_{(i,j)\in\mathcal{P}_{\text{live}}} n_{ij}(t) < \sqrt{t}/|\mathcal{P}_{\text{live}}|$ and, if so, gives the batch to a least-sampled live pair. This keeps every available pair under sampling while spending a vanishing fraction of the effort asymptotically. After selecting a pair the algorithm casts b votes, updates N_t and t, and refits warm-started. A pair that can supply no more votes leaves $\mathcal{P}_{\text{live}}$; if none remains before the stopping condition is met, the procedure returns EXHAUSTED. On a finite archive this is the common outcome, and Section 4 reports it as such rather than as a time.

Theorem 3.3 (Upper bound on the certification time)**.** *Under Assumptions 2.6–2.7, the procedure of this section, stopping on Λ_A at the boundary of Definition 3.1, satisfies*

$$\limsup_{\delta\to 0} \frac{\mathbb{E}[\tau_\delta]}{\ln(1/\delta)} \;\le\; T^\dagger(\theta) \;:=\; \max_{A\in\mathcal{W}} \frac{1}{\psi_-(\bar{\Delta}_A)\, S_A(w^\star;\theta)}, \qquad \bar{\Delta}_A = \limsup_{t\to\infty} \Delta\big(z_{e,t}^A\big),$$

where $z_{e,t}^A$ is the exact minimiser of Λ_A at time t, ψ_- is as in Lemma F.9, and $w^\star$ attains $\Gamma(\theta)$.

Since $\psi_- \le 1$ we have $T^\dagger \ge T^*(\theta)$, with equality exactly when $\bar{\Delta}_A = 0$ at the maximising region. That case is not approached here: under a fixed design with every pair weight bounded away from zero the contrast width tends to a nonzero constant set by the margin rather than by the sample size (Remark F.15), and the measured penalty is $1.15\times$ on the top-20 fit and $1.27\times$ at $K = 16$. The per-region form matters: taking ψ_- at a single width uniform over regions would give $4.08\times$ and $3.88\times$ instead. Appendix E proves the theorem and tabulates the constants.

4 EXPERIMENTAL ANALYSIS

We evaluate whether the proposed transportation costs diagnose unsupported leaderboard claims and whether certificate-directed routing reduces the number of votes required for certification. We use both recorded Chatbot Arena votes and controlled Bradley–Terry simulations. All experimental details are in Appendix G.

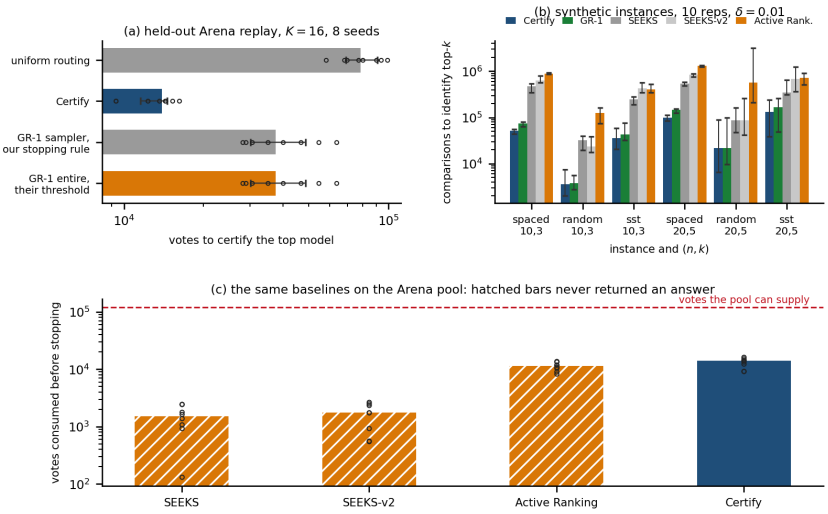

Figure 3: **Against the concurrent optimal algorithm and its non-parametric baselines.** *(a)* Top-1 on the held-out Arena replay. Under a common stopping rule, routing by witness gradient certifies in a median 13,949 votes against 37,400 for the GR-1 sampler and 78,447 for uniform, a paired allocation advantage of 3.24×. *(b)* All five rules on the three synthetic instances of (Goldberger and Rudi, 2026). GR-1 is faster than the three non-parametric baselines in every cell by 2–6×, which is the ordering that paper reports and a check on our port of it; Certify is faster than GR-1 on five of the six cells, by a median 1.24× and at most 1.46×, and level on random 20,5 (0.99×). *(c)* The same baselines on the Arena pool, against CERTIFY for scale. Hatched bars never returned an answer because they exhaust the supply for one *particular* pair they insist on comparing.

Auditing a published leaderboard. The August 2024 Chatbot Arena release contains 1,799,991 records, including 1,157,137 anonymous battles with a decisive winner across 129 models. We fit Bradley–Terry scores to the top $K \in \{20, 40, 64\}$ models by battle count and ask whether the collected votes certify the fitted ordering. These subsets contain respectively 355,507; 668,508; and 900,165 within-set battles. Figure 1 provides an overview and Table 4 shows a sharp difference between global and local resolution. At $K = 64$, the data certify 92.3% of all observed pairwise orderings but only 13.5% of consecutive steps under simultaneous control. Even without multiplicity correction, fewer than half of the consecutive steps are certified. This is compatible with the high split-half Kendall correlations (0.978, 0.971, and 0.967 for $K = 20, 40, 64$): the broad ordering is stable because most models are far apart, whereas one-position differences depend on the smallest and least-resolved gaps. Thus, the leaderboard is globally reproducible but locally underdetermined. As a model-free comparison, exact binomial tests using only direct votes certify 17.3% of consecutive top-64 pairs at level .05, and none after Bonferroni correction. The Bradley–Terry analysis certifies 46.2% at the per-comparison level by sharing information through the comparison graph. Details are described in Appendix G.2.

Best-model identification on recorded votes. Certificate-directed routing, CERTIFY, uses 31% of the votes required by uniform routing under the same GLR stopping rule (Figure 2, Table 5). The gain is explained by where the methods sample. Uncertainty-only rules devote votes to poorly observed but easy comparisons, which are common in organically collected data because obviously mismatched systems are rarely compared. CERTIFY additionally weights uncertainty by its relevance to the nearest wrong answer and therefore concentrates on the close competitors that block certification. Details are described in Appendices G.2 and G.3.

Comparison with the concurrent optimal algorithm. Goldberger and Rudi (2026) proposed the first asymptotically optimal algorithm for identifying top-k arms using pairwise comparisons. We implement their Algorithm 1 without modification and validate the port on the synthetic instances of that paper, where it is fastest in every cell by 2–6× over the non-parametric baselines SEEKS, SEEKS-V2 and ACTIVE RANKING, reproducing the ordering they report. On the held-out Arena replay under a common stopping rule, routing by witness gradient certifies the top model in a median 7,506 votes, against 17,090 for their sampler and 32,170 for uniform; paired seed by seed the

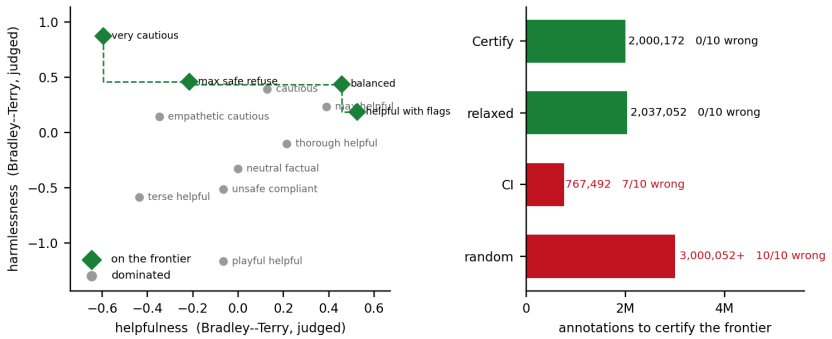

Figure 4: **Certificate-directed annotation on a measured helpful–harmless tradeoff.** *Left:* twelve response policies: system-prompt variants of `google/gemma-4-31B-it` spanning `unsafe_compliant` to `max_safe_refuse` placed by Bradley–Terry scores fitted separately on each axis from real pairwise judgments of their answers to HH-RLHF prompts, order-randomised to control position bias.

allocation advantage is $2.20\times$ (range 1.66–3.76). Run entire under its own threshold, their algorithm never stops and consumes the whole replay half: at $t \approx 17{,}000$ their proved threshold is 62.95 where the boundary our runs use is 9.16, so an end-to-end comparison at moderate δ measures thresholds rather than samplers. We therefore compare allocation under a common rule. Their optimality is a $\delta \to 0$ statement, and this is evidence about the practical regime rather than a counterexample to it. See Appendices G.8 and G.9 and Figure 3.

Two-attribute preference collection. To test genuinely multidimensional alternatives, we construct twelve candidates on a helpfulness–harmlessness tradeoff and collect attribute-specific comparisons in Figure 4. The true frontier contains four candidates. Across 25 seeds, CERTIFY requires 287,252 votes at $\delta = .05$, compared with 301,972 for RELAXED and 1,067,892 for random routing. Thus, certificate-directed sampling is $3.7\times$ cheaper than random sampling, while exact joint pricing improves over the relaxation by approximately 5%. This is the setting in which the joint geometry matters: different dominators must be escaped along different attribute axes. We also repeat this experiment with twelve response policies instantiated by a real language model and judged on helpful and harmless prompts. At $\delta = .05$, CERTIFY and RELAXED stop with the correct frontier in all ten seeds, at median costs of 2,000,172 and 2,037,052 votes. Random routing is censored at three million votes in all seeds. ELO-CI uses fewer votes but succeeds in only three of ten runs: three runs are censored and four of its seven early stops are false. Thus, its lower vote count does not provide anytime-valid certification on this harder, small-gap instance. Details are in Appendix G.4.

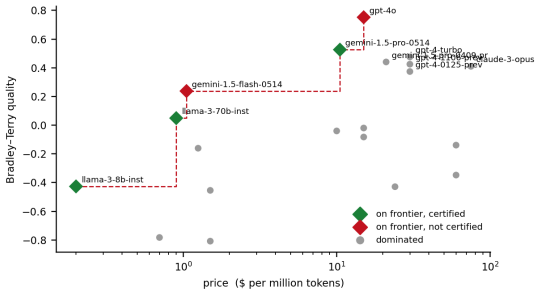

Figure 5: **Arena top-20 quality–cost frontier.** Five of twenty models are undominated (dashed staircase). Green: membership certified against $\beta = 91$. Red: not. `gpt-4o` reaches 84% of what its membership requires and `gemini-1.5-flash` 71%.

Frontier identification. We next consider a quality–cost frontier among the top-20 Arena models recovered as Figure 5. In Bradley–Terry replay, the true frontier contains five models. Routing reduces the simulated certification cost by approximately $5.5\times$ relative to uniform sampling. Exact joint pricing and the weakest-link relaxation are indistinguishable on this instance: their median counts differ by less than 1%. The dominators are nearly aligned in the quality–cost plane, so one displacement almost escapes them all and the joint price is close to its cheapest component. Details are described in Appendices G.4, G.5 and G.7.

AI USE STATEMENT

We used generative AI tools (Claude) to edit manuscript, write and review the code, to draft and check several appendix derivations, and to assist with running and monitoring experiments. The research questions, the method, and the experimental design are the authors'. All derivations were verified by the authors, all numbers are generated from logged results by scripts the authors reviewed, and the authors take full responsibility for the content.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

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: 10.2307/2334029.

Dallas Card, Peter Henderson, Urvashi Khandelwal, Robin Jia, Kyle Mahowald, and Dan Jurafsky. With little power comes great responsibility. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9263–9274. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.emnlp-main.745.

Ivi Chatzi, Eleni Straitouri, Suhas Thejaswi, and Manuel Gomez Rodriguez. Prediction-powered ranking of large language models. In *Advances in Neural Information Processing Systems*, volume 37, pages 113096–113133, 2024. doi: 10.52202/079017-3594.

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*, pages 8359–8388, 2024.

Nirjhar Das, Souradip Chakraborty, Aldo Pacchiano, and Sayak Ray Chowdhury. Active preference optimization for sample efficient RLHF. In *Machine Learning and Knowledge Discovery in Databases. Research Track (ECML PKDD 2025)*, Lecture Notes in Computer Science, pages 96–112. Springer Nature Switzerland, 2025a. doi: 10.1007/978-3-032-06096-9_6.

Udvas Das, Apurv Shukla, and Debabrota Basu. FraPPE: Fast and efficient preference-based pure exploration. In *Advances in Neural Information Processing Systems*, volume 38, pages 9225–9271, 2025b. doi: 10.52202/085713-0312.

Rémy Degenne and Wouter M. Koolen. Pure exploration with multiple correct answers. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Mostafa Dehghani, Yi Tay, Alexey A. Gritsenko, Zhe Zhao, Neil Houlsby, Fernando Diaz, Donald Metzler, and Oriol Vinyals. The benchmark lottery, 2021. URL https://arxiv.org/abs/2107.07002.

Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 998–1027. PMLR, 2016. URL https://proceedings.mlr.press/v49/garivier16a.html.

Motti Goldberger and Nils Rudi. Optimal top-k identification from pairwise comparisons, 2026. URL https://arxiv.org/abs/2607.08979.

Reinhard Heckel, Max Simchowitz, Kannan Ramchandran, and Martin Wainwright. Approximate ranking from pairwise comparisons. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1057–1066, 2018.

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.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021. doi: 10.1214/20-AOS1991.

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, pages 26791–26803, 2022. doi: 10.52202/068431-1943.

Emilie Kaufmann and Wouter M. Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44, 2021.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Adaptive algorithms for relaxed Pareto set identification. In *Advances in Neural Information Processing Systems*, volume 36, pages 35190–35201, 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*, pages 2548–2556, 2024.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Constrained Pareto set identification with bandit feedback. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 31342–31378, 2025.

Jiachun Li, David Simchi-Levi, and Will Wei Sun. Low rank for rank: Uncertainty-aware task-specific LLM ranking under sparse pairwise comparisons, 2026. URL https://arxiv.org/abs/2605.29395.

Zhaoqi Li, Kevin Jamieson, and Lalit Jain. Optimal exploration is no harder than Thompson sampling. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 1684–1692, 2024. URL https://proceedings.mlr.press/v238/li24h.html.

Viraj Mehta, Syrine Belakaria, Vikramjeet Das, Ojash Neopane, Yijia Dai, Ilija Bogunovic, Barbara Engelhardt, Stefano Ermon, Jeff Schneider, and Willie Neiswanger. Sample efficient preference alignment in LLMs via active exploration, 2023. URL https://arxiv.org/abs/2312.00267.

Bitya Neuhof and Yuval Benjamini. Quantifying ranking uncertainty in LLM benchmarks, 2026. URL https://arxiv.org/abs/2607.16259.

Hosna Oyarhoseini, Jimmy Lin, and Amir-Hossein Karimi. A unified perturbation framework for analyzing leaderboard stability and manipulation, 2026. URL https://arxiv.org/abs/2605.15761.

Yotam Perlitz, Elron Bandel, Ariel Gera, Ofir Arviv, Liat Ein-Dor, Eyal Shnarch, Noam Slonim, Michal Shmueli-Scheuer, and Leshem Choshen. Efficient benchmarking (of language models). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2519–2536. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.naacl-long.139.

Wenbo Ren, Jia Liu, and Ness Shroff. The sample complexity of best-k items selection from pairwise comparisons. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 8051–8072, 2020.

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, Sivaraman Balakrishnan, Joseph Bradley, Abhay Parekh, Kannan Ramchandran, and Martin J. Wainwright. Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. *Journal of Machine Learning Research*, 17(58):1–47, 2016.

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

Po-An Wang, Ruo-Chun Tzeng, and Alexandre Proutiere. Fast pure exploration via Frank-Wolfe. In *Advances in Neural Information Processing Systems*, volume 34, pages 5810–5821, 2021.

Hongli Zhou, Hui Huang, Ziqing Zhao, Lvyuan Han, Huicheng Wang, Kehai Chen, Muyun Yang, Wei Bao, Jian Dong, Bing Xu, Conghui Zhu, Hailong Cao, and Tiejun Zhao. Lost in benchmarks? Rethinking large language model benchmarking with item response theory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 35085–35093, 2026. doi: 10.1609/aaai.v40i41.40814.

A NOTATION

Notation conventions. Unprimed parameters denote the fitted or data-generating score vector, while θ' denotes an alternative parameter. A calligraphic symbol denotes a set or collection; a Roman capital denotes a single convex witness region. The exact likelihood price Λ_P is used for certified stopping, whereas the quadratic price S_P and its gradient are used for routing.

Table 1: Notation used throughout the paper.

Symbol	Meaning
Models, comparisons, and observations	
K, $[K]$	Number of models and model index set $\{1, \dots, K\}$.
$\theta \in \mathbb{R}^K$	Bradley–Terry score vector; scores are identifiable only up to an additive constant.
$V = \{\vartheta : \mathbf{1}^\top \vartheta = 0\}$	Centered contrast subspace used to identify the score vector.
Θ_R	Bounded parameter set $\{\vartheta \in V : \|\vartheta\|_\infty \le R\}$.
$c = (c_i)_{i\in[K]}$	Model costs used in the quality–cost frontier.
$\mathcal{P}$	Set of unordered comparison pairs, $\{(i,j) : i < j\}$; $\|\mathcal{P}\| = \binom{K}{2}$.
$A_t = (i,j)$	Pair selected for comparison at round t.
Z_t	Bernoulli outcome of comparison A_t; for $A_t = (i,j)$, $Z_t = 1$ denotes a win by model i.
$\mathcal{F}_t$	Filtration generated by the selected pairs and observed outcomes through round t.
$N_t = (n_{ij}(t))$	Pairwise comparison-count vector after t votes.
W_{ij}	Number of observed wins by model i over model j; $n_{ij} = W_{ij} + W_{ji}$.
$w \in \Delta_{\mathcal{P}}$	Normalized allocation over comparison pairs; $\Delta_{\mathcal{P}}$ is the probability simplex on $\mathcal{P}$.
$\widehat{w}_t = N_t/t$	Empirical comparison allocation after t votes.
$\mathcal{P}_{\text{live}}$	Set of pairs from which additional votes can still be collected.
Bradley–Terry model and information geometry	
$\sigma(x) = (1 + e^{-x})^{-1}$	Logistic link; $\Pr_\theta(i \succ j) = \sigma(\theta_i - \theta_j)$.
$\eta_{ij}(\theta) = \theta_i - \theta_j$	Bradley–Terry log-odds contrast for pair (i,j).
$\ell_N(\theta)$	Bradley–Terry log-likelihood based on comparison counts and outcomes summarized by N.
$\widehat{\theta}_t$	Bradley–Terry maximum-likelihood estimate after t votes.
e_i	The ith standard basis vector in $\mathbb{R}^K$.
$L(N;\vartheta)$	Information-weighted graph Laplacian, $\sum_{i<j} n_{ij}\sigma'(\vartheta_i - \vartheta_j)(e_i - e_j)(e_i - e_j)^\top$.
$L(N)^+$	Moore–Penrose pseudoinverse of the information Laplacian.
$R_{ij}(N)$	Effective resistance $(e_i - e_j)^\top L(N)^+ (e_i - e_j)$.
$\|x\|_L^2 = x^\top L x$	Squared seminorm induced by a positive-semidefinite matrix L.
$\mathrm{kl}(p,q)$	Binary Kullback–Leibler divergence between Bernoulli distributions with means p and q.
$d(w;\theta,\theta')$	KL transportation cost of moving from θ to θ' under allocation w.
$T^*_{\mathrm{KL}}(\theta)$	Instance-wise asymptotic lower-bound constant based on the exact KL transportation cost.
Claims, alternatives, and witnesses	
$\mathcal{C}$	Finite answer space of the claim under study.
$\mathrm{ans} : \Theta \to \mathcal{C}$	Answer map, such as the best model, top-k set, or frontier.
$\widehat{a}_t = \mathrm{ans}(\widehat{\theta}_t)$	Candidate answer after t votes.
$\mathrm{Alt}(\theta)$	Parameters producing an answer different from $\mathrm{ans}(\theta)$.
$\mathcal{U}(\theta)$	Interior of the parameter cell on which the answer equals $\mathrm{ans}(\theta)$.

Continued on next page

Table 1 continued

Symbol	Meaning
$\mathcal{W}(\theta)$	Closed witness set $\overline{\mathrm{Alt}(\theta)} = \Theta \setminus \mathcal{U}(\theta)$.
P	A closed convex witness region (or a piece in a decomposition/cover of the witness set).
$\mathfrak{C}(\theta)$	Finite collection of convex regions covering $\mathcal{W}(\theta)$.
A_j	Top-1 witness region in which challenger j ties or exceeds the current leader.
A_{ij}	Top-k witness region in which an excluded model j ties or exceeds included model i.
X_{ij}	Frontier-exit region in which model j causes frontier model i to leave the frontier.
E_m	Frontier-entry region in which currently dominated model m enters the frontier.
D_m	Set of models that currently dominate model m.
F	Current quality–cost or multi-attribute frontier.
Evidence statistics and stopping	
$\Lambda(\theta'; N)$	Likelihood deficit $\ell_N(\widehat{\theta}) - \ell_N(\theta')$.
$\Lambda_P(N)$	Exact generalized likelihood-ratio price $\inf_{\theta' \in P} \Lambda(\theta'; N)$.
$S_P(N)$	Quadratic price $\frac{1}{2} \inf_{\theta' \in P} \|\theta' - \widehat{\theta}\|^2_{L(N)}$, used for routing.
z_P	Displacement from $\widehat{\theta}$ to a minimizer of the quadratic projection defining S_P.
$\Delta(z)$	Contrast width $\max_{i<j} \|z_i - z_j\|$.
$\psi_-(x), \psi_+(x)$	Scalar comparison functions that bound the ratio between Λ_P and S_P.
$\beta(t, \delta)$	Proved time-uniform stopping boundary for the exact likelihood statistic.
$\beta_h(t, \delta)$	Proposed boundary indexed by active-constraint dimension h.
δ	Target probability of returning an incorrect claim.
τ	First time at which every relevant witness region crosses its stopping boundary.
h_P	Number of linearly independent constraints active at the projection onto region P.
m_P	Number of linear inequalities defining region P.
Adaptive routing and asymptotic quantities	
S_t^*	Smallest quadratic witness price at time t, $\min_{P \in \mathfrak{C}_t} S_P$.
$\mathcal{B}_t$	Set of near-binding witness regions whose prices are within the shrinking window of S_t^*.
$G_{ij}(t)$	Routing score: estimated marginal increase in the near-binding quadratic prices from another vote on pair (i, j).
ρ	Exponent controlling the shrinking near-binding window; the algorithm uses $\rho = 0.9$.
b	Number of votes collected in each selected batch.
$\Phi(w; \vartheta)$	Worst-region quadratic information rate $\min_{P \in \mathcal{W}} S_P(w; \vartheta)$.
$\Gamma(\vartheta)$	Optimal quadratic information rate $\max_{w \in \Delta_{\mathcal{P}}} \Phi(w; \vartheta)$.
$T^*(\theta) = \Gamma(\theta)^{-1}$	Quadratic asymptotic certification-time constant.

B Technical Details: Witnesses, prices, and costs

Definition B.1 (Deficit, price, and binding witness)**.** *The* deficit *of a witness θ' is*

$$\Lambda(\theta'; N) = \ell_N(\hat{\theta}) - \ell_N(\theta') \geq 0.$$

It is the log-likelihood lost by replacing the fit $\hat{\theta}$ with θ'; equivalently, the votes are $e^{\Lambda(\theta'; N)}$ times more likely under $\hat{\theta}$. For $B \subseteq \Theta$, its price *is*

$$\Lambda_B(N) = \inf_{\theta' \in B} \Lambda(\theta'; N).$$

The price of a claim is $\Lambda_{\mathcal{W}}(N)$, the smallest deficit among its witnesses. A minimiser is a binding witness.

Lemma B.2 (Attainment)**.** *If $L(N) \succ 0$ on $\mathbf{1}^{\perp}$, then the infimum defining Λ_B is attained for every nonempty closed B, and the minimiser is unique when B is convex.*

Proof. On $\mathbf{1}^{\perp}$, $\Lambda(\cdot; N)$ is continuous, strictly convex, and coercive because ℓ_N is strictly concave with a unique maximiser. A continuous coercive function attains its infimum on every nonempty closed set, uniquely when both the set is convex and the function is strictly convex. □

Why witnesses must be split. The set $\mathcal{W}(\hat{\theta})$ is nonconvex, so $\Lambda_{\mathcal{W}}$ is not a single projection. Nor can $\mathcal{W}$ be replaced by its convex hull. For top-1 with $\hat{\theta} = (2, -1, -1)$, the witnesses $(2, 2.1, -4.1)$ and $(2, -4.1, 2.1)$ place different rivals on top, but their average is $\hat{\theta}$. Thus, the convex hull of $\mathcal{W}$ contains the fit and has price zero.

Every claim considered here depends only on the order of the scores on two orders for the two-attribute frontier. The hyperplanes $\{\theta_i = \theta_j\}$ partition Θ into $K!$ closed ordering chambers, and ans is constant on the interior of each chamber. Consequently, every witness set is a union of convex chambers. This generic decomposition is computationally useless: for top-1 with $K = 20$, it contains all but $(K-1)!$ of the chambers, approximately 2.3×10^{18} pieces. What is needed is a small collection of pieces, each described by few inequalities, because computing the price requires one convex solve per piece.

Proposition B.3 (Small decompositions)**.** *For each claim in Table 2 and every $\hat{\theta}$ at which the claim is strict, $\mathcal{W}(\hat{\theta})$ is exactly the union of the listed closed convex pieces.*

Table 2: Decompositions of the witness set. Here F is the current frontier and c denotes API cost. In the two-attribute setting, j dominates i when it is strictly better on both axes.

claim	pieces	number	ineq. each
top-1, leader a	$A_j = \{\theta'_j \geq \theta'_a\}$, $j \neq a$	$K-1$	1
top-k, set S	$A_{aj} = \{\theta'_j \geq \theta'_a\}$, $a \in S$, $j \notin S$	$k(K-k)$	1
quality–cost frontier	exits $X_{ij} = \{\theta'_j \geq \theta'_i\}$, $i \in F$, $j \neq i$, $c_j \leq c_i$; entries $E_m = \bigcap_{j \neq m,\, c_j \leq c_m} \{\theta'_m \geq \theta'_j\}$, $m \notin F$	$\leq \lvert F\rvert(K-1) + K - \lvert F\rvert$	1; $\leq K-1$
two-attribute frontier	exits $\{h'_j \geq h'_i\} \cap \{s'_j \geq s'_i\}$, $i \in F$, $j \neq i$; entries $E^{\alpha}_m = \bigcap_{j \neq m} \{m$ is at least as good as j on axis $\alpha_j\}$, $\alpha \in \{h, s\}^{[K] \setminus \{m\}}$	$\leq \lvert F\rvert(K-1) + (K - \lvert F\rvert)2^{K-1}$	2; $K-1$
any order-based claim	closed ordering chambers	$\leq K!$	$K-1$

Corollary B.4 (Price as a minimum of convex problems)**.** *Over the pieces in Proposition B.3,*

$$\Lambda_{\mathcal{W}}(N) = \min_{P} \Lambda_P(N).$$

The binding witness of the cheapest piece is a binding witness of the claim.

Proof. An infimum over a finite union is the minimum of the component infima. □

By Lemma B.2, each Λ_P minimises a strictly convex function over a polyhedron and has a unique minimiser. For a single half-space, its second-order value is $\mathrm{gap}^2/(2R_{ij})$. The third column of Table 2 is therefore the number of convex problems solved at each stopping check. For top-1, the small decomposition reduces this number from approximately 2.3×10^{18} to $K-1$. The only expensive case is a two-attribute entry, which has 2^{K-1} pieces.

Remark B.5 (Relation to Goldberger and Rudi (Goldberger and Rudi, 2026))**.** *Their boundary-pair set $B(\theta)$ is exactly the top-k row of Table 2. The frontier rows extend this construction to claims that are not top-k sets.*

C Detailed Pseudocodes

Algorithm 2 COVER$(\text{ans}, \hat{\theta})$; pieces covering the witness set

Ensure: a finite collection $\mathcal{C}$ of closed convex sets satisfying $\mathcal{W}(\hat{\theta}) \subseteq \bigcup \mathcal{C}$

if the claim is **top-**1 **then**
  $a \leftarrow \arg\max_i \hat{\theta}_i$
  return $\big\{\{\theta' : \theta'_j \geq \theta'_a\}\big\}_{j \neq a}$
else if the claim is **top-**k, with set S **then**
  return $\big\{\{\theta' : \theta'_j \geq \theta'_a\}\big\}_{a \in S,\, j \notin S}$
else if the claim is a **quality–cost frontier then**
  $F \leftarrow \{i : \nexists j,\ \hat{\theta}_j > \hat{\theta}_i \text{ and } c_j \leq c_i\}$
  $D_m \leftarrow \{j : \hat{\theta}_j > \hat{\theta}_m,\ c_j \leq c_m\}$ for $m \notin F$
  return the exits $\{\theta' : \theta'_j \geq \theta'_i\}$ for $i \in F$, $j \neq i$, $c_j \leq c_i$,
    and the pruned entries $E_m^D = \bigcap_{j \in D_m} \{\theta' : \theta'_m \geq \theta'_j\}$ for $m \notin F$
else if the claim is a **two-attribute frontier then**
  $F \leftarrow$ Pareto frontier of $(\hat{h}, \hat{s})$
  $D_m \leftarrow$ the models dominating m on both axes
  return the exits $\{h'_j \geq h'_i\} \cap \{s'_j \geq s'_i\}$ for $i \in F$, $j \neq i$,
    and, for each $m \notin F$ and $\alpha \in \{h, s\}^{D_m}$, the entry piece $\bigcap_{j \in D_m} \{m \text{ is at least as good as } j \text{ on axis } \alpha_j\}$

For top-1 and top-k, the returned pieces give exact decompositions of $\mathcal{W}(\hat{\theta})$ into $K-1$ and $k(K-k)$ half-spaces. For a frontier entry E_m, the algorithm retains only the constraints corresponding to models that currently dominate m:

$$E_m^D = \bigcap_{j \in D_m} \{\theta'_m \geq \theta'_j\} \supseteq E_m.$$

The resulting collection may be a strict cover: it can contain parameters at which m overtakes its current dominators but is itself overtaken by a previously lower-ranked, cheaper model. This enlargement is safe by Lemma 2.5, because a larger set has a smaller price and can only delay stopping. Exactness is recovered by checking the minimiser after each solve and, whenever it violates a dropped inequality, solving the full entry set instead. For a two-attribute frontier, the pruning reduces an entry from 2^{K-1} pieces to $2^{|D_m|}$.

Algorithm 3 QUADPRICE$(P, L, \hat{\theta})$; projection by active-set enumeration

Require: $P = \bigcap_{k \in [m_P]} \{\theta' : \langle y_k, \theta' - \hat{\theta} \rangle \geq g_k\}$

Ensure: $S_P = \frac{1}{2} \min_{\theta' \in P} \|\theta' - \hat{\theta}\|_L^2$ and the displacement z_P from $\hat{\theta}$ to the minimiser

$S^\star \leftarrow \infty$; $Y \leftarrow [y_1; \ldots; y_{m_P}]$; $L^+ \leftarrow (L + \mathbf{1}\mathbf{1}^\top / K)^{-1}$
for all $\emptyset \neq \mathcal{A} \subseteq [m_P]$ **do** ▷ candidate active sets
  solve $Y_\mathcal{A} L^+ Y_\mathcal{A}^\top \lambda = g_\mathcal{A}$ **(skip if singular)**
  if $\lambda \not\geq 0$ **then continue** ▷ dual feasibility
  $z \leftarrow L^+ Y_\mathcal{A}^\top \lambda$
  if $\langle y_k, z \rangle < g_k$ for some $k \notin \mathcal{A}$ **then continue** ▷ primal feasibility
  if $\frac{1}{2} g_\mathcal{A}^\top \lambda < S^\star$ **then** $S^\star \leftarrow \frac{1}{2} g_\mathcal{A}^\top \lambda$; $z^\star \leftarrow z$
return $(S^\star, z^\star)$

For every top-1 and top-k piece, $m_P = 1$, so the loop has one term and $S_P = \text{gap}^2/(2R_{ij})$. Only frontier entries have $m_P > 1$. For two-attribute pieces, L is block diagonal across the two independent axes, so each projection is the sum of one helpfulness and one harmlessness projection.

Algorithm 4 EXACTPRICE$(P, N, \hat{\theta})$; exact price for stopping

Require: P as in Algorithm 3; counts N
Ensure: $\Lambda_P(N) = \min_{\theta' \in P}[\ell_N(\hat{\theta}) - \ell_N(\theta')]$
1: $\Lambda^\star \leftarrow \infty$
2: **for all** $\emptyset \neq \mathcal{A} \subseteq [m_P]$ **do**
3: minimise $-\ell_N(\theta')$ by Newton subject to $\langle y_k, \theta' - \hat{\theta}\rangle = g_k$, $k \in \mathcal{A}$; call the minimiser $\theta^{\mathcal{A}}$
4: **if** a multiplier is negative, or $\langle y_k, \theta^{\mathcal{A}} - \hat{\theta}\rangle < g_k$ for some $k \notin \mathcal{A}$, **then continue**
5: $\Lambda^\star \leftarrow \min\{\Lambda^\star, \ell_N(\hat{\theta}) - \ell_N(\theta^{\mathcal{A}})\}$
6: **return** $\Lambda^\star$

This is the same active-set enumeration as QUADPRICE, with the exact log-likelihood replacing its quadratic approximation. The KKT test is valid because $-\ell_N$ is convex and the constraints are linear. For a single half-space $\{\theta'_j \geq \theta'_a\}$ with $\hat{\theta}_a > \hat{\theta}_j$, the minimiser lies on the boundary, so the algorithm requires one Newton solve under $\theta'_j = \theta'_a$.

D TECHNICAL RESULTS

D.1 PROOF OF PROPOSITION B.3

Proof. Each piece P is the closure of its nonempty strict version P°. It therefore suffices to show that (a) $P^\circ \subseteq \mathrm{Alt}(\hat{\theta})$ for every piece, implying $P = \overline{P^\circ} \subseteq \overline{\mathrm{Alt}} = \mathcal{W}$, and (b) $\mathrm{Alt}(\hat{\theta}) \subseteq \bigcup P$, implying $\mathcal{W} = \overline{\mathrm{Alt}} \subseteq \bigcup P$ because a finite union of closed sets is closed.

Top-1. If $\theta'_j > \theta'_a$, then a is not the strict maximum, so $\mathrm{ans}(\theta') \neq a$. Conversely, if $\mathrm{ans}(\theta') \neq a$, then some $j \neq a$ satisfies $\theta'_j \geq \theta'_a$.

Top-k. If some $j \notin S$ strictly exceeds some $a \in S$, then S is not the top-k set. Conversely, if S is not the strict top-k set at θ', then not every member of S strictly exceeds every non-member.

Quality–cost frontier. In X°_{ij}, model j is at least as cheap as i and strictly better, so $i \in F$ leaves the frontier. In E°_m, model m strictly exceeds every model at least as cheap, so $m \notin F$ joins the frontier. Conversely, if the frontier at θ' differs from F, then either some $i \in F$ is dominated by a model j satisfying $c_j \leq c_i$ and $\theta'_j > \theta'_i$, in which case $\theta' \in X_{ij}$, or some $m \notin F$ is undominated. In the latter case, $\theta'_m \geq \theta'_j$ for every j at least as cheap as m, so $\theta' \in E_m$.

Two-attribute frontier. The exit argument is identical, with dominance required on both axes. A model $m \notin F$ joins the frontier exactly when no j dominates it, or equivalently, when for every $j \neq m$, m is at least as good as j on some axis. This is an intersection over j of unions of two half-spaces. Distributing the intersection over the unions yields the union, over assignments α of an axis to each j, of the convex pieces E^α_m. Their strict versions lie in Alt, establishing (a). □

D.2 PROOF OF LEMMA 2.5

Proof. Because $\bigcup \mathcal{C} \supseteq \mathcal{W}$, taking the infimum over the cover gives $\Lambda_{\mathcal{C}} \leq \Lambda_{\mathcal{W}}$. If the minimiser $z^\star$ of the cheapest $C \in \mathcal{C}$ lies in $\mathcal{W}$, then $\Lambda_{\mathcal{W}} \leq \Lambda(z^\star; N) = \Lambda_{\mathcal{C}}$, proving equality. □

E SAMPLE COMPLEXITY ANALYSIS

Throughout, $m = |\mathcal{P}| = \binom{K}{2}$, $\Theta_R = \{\vartheta \in V : \|\vartheta\|_\infty \leq R\}$, $a = \sigma'(2R) > 0$, and $\hat{w}_t = N_t/t$ is the empirical design. Write

$$\Phi(w; \vartheta) = \min_{A \in \mathcal{W}} S_A(w; \vartheta), \qquad \Gamma(\vartheta) = \max_{w \in \Delta_{\mathcal{P}}} \Phi(w; \vartheta), \qquad T^*(\theta) = \Gamma(\theta)^{-1}.$$

Theorem E.1 (Upper bound on the certification time)**.** *Under Assumptions 2.6–2.7, the procedure of Section 3, stopping on Λ_A at the boundary of Theorem 3.2, satisfies*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\ln(1/\delta)} \;\leq\; T^\dagger(\theta) \;:=\; \max_{A \in \mathcal{W}} \frac{1}{\psi_-(\bar{\Delta}_A)\, S_A(w^\star; \theta)}, \qquad \bar{\Delta}_A = \limsup_{t \to \infty} \Delta\big(z^A_{e,t}\big),$$

where $z^A_{e,t}$ is the exact minimiser of Λ_A at time t, ψ_- is as in Lemma F.9 and $w^\star$ attains $\Gamma(\theta)$. Since $\psi_- \le 1$, $T^\dagger \ge T^*(\theta)$, with equality iff the surrogate is exact at the binding region.

Remark E.2 (Why per region, and what it is worth)**.** *The cruder statement $T^*(\theta)/\psi_-(\max_A \bar\Delta_A)$ also follows from the proof, by applying the sandwich uniformly before taking the minimum. It is badly loose. The maximum contrast width is attained at the* easy *regions, where the displacement needed to reach the boundary is large, on the Arena top-*20*, $\max_A D^\star_A = 6.99$ and $\psi_- = 0.245$, an inflation of $4.08\times$, but those regions also have enormous S_A and never bind. Taking the maximum region by region, after the sandwich rather than before, gives $1.15\times$ on the same instance, a factor of 3.5 tighter, and the binding region turns out to be the one with the* smallest *contrast width. The revision costs nothing: it reorders two steps in the assembly.*

Remark E.3 (The ψ_- factor cannot be removed by stopping on S_A)**.** *It is tempting to drop the factor by stopping on the quadratic price directly, since S_A is what the routing already computes. That rule is* not *δ-correct. Theorem 3.2 rests on $\Lambda_A \le \ell_t(\hat\theta_t) - \ell_t(\theta)$ whenever $\theta \in A$, an identity for the exact log-GLR that has no analogue for the surrogate: by Theorem F.10, $S_A \in [\Lambda_A/\psi_+, \Lambda_A/\psi_-]$, and $\psi_- < 1$, so S_A may* exceed *Λ_A, by up to 7% at $\bar\Delta = 0.22$ and 12% at $\bar\Delta = 0.35$, the margins these experiments run at. Testing $S_A \ge \beta$ can therefore fire while Λ_A is only $\psi_-\beta < \beta$, before the martingale argument licenses it.*

The valid version inflates the threshold, $S_A \ge \beta/\psi_-(D^\star)$ (Corollary F.16) and then the crossing time satisfies $t\,\Phi(\hat w_t) \ge \beta/\psi_-$, so the asymptotic constant is again $T^*(\theta)/\psi_-$. The two routes cost the same, which is the point: the factor is not an artefact of where the test is placed but the price of routing on a surrogate while demanding a guarantee on the exact statistic*. Removing it requires routing on Λ_A itself which also makes T^* the Kullback–Leibler characteristic time and so closes the gap with the lower bound. See Remark E.13.*

Two things about the statement. It is an *upper* bound; what a matching lower bound would require is set out in Remark E.13. And the factor $\psi_-(\bar\Delta)^{-1}$ is the price of stopping on the exact log-GLR while routing on the quadratic surrogate: by Theorem F.10 the two agree only within a bracket fixed by the margin, and Remark F.15 shows that bracket does not close as $t \to \infty$. It is $1+\bar\Delta/3+O(\bar\Delta^2)$, so it is mild exactly when the binding margins are small.

The proof has five steps: forcing gives a deterministic floor on every pair count (Lemma E.4); that floor plus a self-normalised bound gives a rate for $\hat\theta_t$ (Lemma E.7) and a summable tail for the time after which the estimate stays accurate (Corollary E.8); the price is regular enough for Frank–Wolfe (Lemma E.10); the design therefore reaches and stays near the maximiser in expected finite time (Lemma E.11); and a deterministic time bounds the stopping rule once the design is good (Lemma E.12).

E.1 Step 1: forcing gives a count floor

Lemma E.4. *Under override (iv), for all $t \ge 16m^2$, $\min_{(i,j)\in\mathcal{P}} n_{ij}(t) \;\ge\; \sqrt{t}/(2m) - 1$.*

Proof. Write $\mu(t) = \min_{ij} n_{ij}(t)$, nondecreasing, and $g(t) = \sqrt{t}/m$. The override is checked at round $t+1$ against $\mu(t) < g(t)$. If round t was not a forcing round then the check failed at that round, so $\mu(t) \ge \mu(t-1) \ge g(t-1) = \sqrt{t-1}/m \ge \sqrt{t}/(2m)$ for $t \ge 2$, and the claim holds.

Otherwise let $s \in \{0,\dots,t-1\}$ be the largest index with round s not forcing ($s = 0$ if all rounds force), so rounds $s+1,\dots,t$ all force. Each forcing round increments a pair currently attaining the minimum. At most m pairs can attain it, and the minimum rises only once every minimiser has been incremented; hence any m consecutive forcing rounds raise μ by at least one, giving $\mu(t) \ge \mu(s) + \lfloor (t-s)/m \rfloor$. As above $\mu(s) \ge \sqrt{s-1}/m$ when $s \ge 1$, and $\mu(s) \ge 0$ always, so

$$\mu(t) \;\ge\; \max\big\{ \tfrac{\sqrt{s-1}}{m}\mathbf{1}\{s \ge 1\},\; \tfrac{t-s}{m} - 1 \big\}.$$

If $s \ge t/2$ the first term gives $\mu(t) \ge \sqrt{t/2-1}/m \ge \sqrt{t}/(2m)$ for $t \ge 4$. If $s < t/2$ the second gives $\mu(t) \ge t/(2m) - 1 \ge \sqrt{t}/(2m) - 1$ since $t \ge \sqrt{t}$. □

Remark E.5. *The floor is $\sqrt{t}/(2m) - 1$, not $\sqrt{t}/m - 1$: the factor 2 is the price of ties at the minimum, which the argument "the minimum can fall below the threshold for at most one round"*

overlooks, since incrementing one of several tied minimisers does not raise the minimum. Only the order $\sqrt{t}$ matters below.

E.2 Step 2: rate for the estimator, and a summable tail

Remark E.6 (Why not Hoeffding on the win frequencies)**.** *Strong convexity of $-\ell_N$ on V is governed by the* smallest *pair count, $L(N;\vartheta) \succeq a\kappa_t K\, I_V$ with $\kappa_t \asymp \sqrt{t}$, whereas perturbing $\hat{p} \to p$ moves the gradient by $\sum_{ij} n_{ij}(\hat{p}_{ij} - p_{ij})(e_i - e_j)$, of norm at most $\sqrt{2}\, t\, \|\hat{p} - p\|_\infty$, governed by the* total *count. The implied Lipschitz constant is therefore $L_0 = \sqrt{2}\, t/(a\kappa_t K) = \Theta(\sqrt{t})$, not $O(1)$, and the resulting exponent $\kappa_t/L_0^2 \asymp t^{-1/2}$ makes the bound vacuous. The mismatch is real, not an artefact of loose constants: forcing guarantees only $\min_{ij} n_{ij} \gtrsim \sqrt{t}$, so $\lambda_2(L)$ may genuinely be $\Theta(\sqrt{t})$ while $\sum_{ij} n_{ij} = t$. We therefore bound the score in its self-normalised norm instead, which is insensitive to this imbalance.*

Lemma E.7 (Estimation rate)**.** *Assume 2.6 and the forcing override, and fix $\lambda > 0$. With probability at least $1 - \delta'$, simultaneously for all $t \geq t_0 := \max\{16m^2, (2\lambda m/(aK))^2\}$,*

$$\|\hat{\theta}_t - \theta\|_2 \;\leq\; C_\star \sqrt{\tilde{\beta}_t(\delta')}\, t^{-1/4}, \qquad \tilde{\beta}_t(\delta') = \tfrac{1}{4}\Big[(K-1)\ln\big(1 + \tfrac{2t}{\lambda}\big) + 2\ln\tfrac{1}{\delta'}\Big],$$

with $C_\star = \big(a\sqrt{aK/(2m)}\big)^{-1}\sqrt{2}$ depending only on K, R.

Proof. Let $S_t = \nabla\ell_t(\theta) = \sum_{s\leq t} \varepsilon_s x_s$ with $\varepsilon_s = Y_s - \sigma(\langle x_s, \theta\rangle)$, and $V_t = \sum_{s\leq t} x_s x_s^\top$, the unweighted Laplacian of the vote graph.

From score to parameter. For $\vartheta \in \Theta_R$ every $|\langle x_s, \vartheta\rangle| \leq 2R$, so $\sigma' \geq a$ and, by the mean value theorem applied coordinatewise to $\vartheta \mapsto \nabla\ell_t(\theta) - \nabla\ell_t(\vartheta)$, there is $\bar{H}_t \succeq aV_t$ on V with $\hat{\theta}_t - \theta = \bar{H}_t^{+} S_t$. Hence $\|\hat{\theta}_t - \theta\|_{V_t} \leq a^{-1}\|S_t\|_{V_t^+}$.

Self-normalised bound. Each ε_s is centred given $\mathcal{F}_{s-1}$ and supported in an interval of length one, hence $\frac{1}{2}$-sub-Gaussian. By Abbasi-Yadkori et al. (2011) on V, with probability $\geq 1 - \delta'$ and simultaneously for all t, $\|S_t\|^2_{(V_t+\lambda I)^{-1}} \leq \tilde{\beta}_t(\delta')$, using $\operatorname{tr} V_t = 2t$ to bound every eigenvalue of V_t by $2t$.

Conditioning. By Lemma E.4, V_t dominates κ_t times the unit-weight complete-graph Laplacian, whose spectrum on V is $\{K\}$; so $\lambda^+_{\min}(V_t) \geq K\kappa_t \geq aK\sqrt{t}/(2m)$ for $t \geq 16m^2$. For $t \geq t_0$ this exceeds λ, so $V_t^+ \preceq 2(V_t + \lambda I)^{-1}$ on V and $\|S_t\|^2_{V_t^+} \leq 2\tilde{\beta}_t(\delta')$.

To the Euclidean norm. $\|v\|_2 \leq \|v\|_{V_t}/\sqrt{\lambda^+_{\min}(V_t)}$, so combining the three displays gives the claim. Everything holds on the single event of the second step, hence simultaneously in t. □

Corollary E.8 (The estimate stays accurate, with a summable tail)**.** *For $\eta > 0$ let $\mathcal{S}_\eta = \inf\{s \geq t_0 :\ \|\hat{\theta}_u - \theta\|_\infty \leq \eta$ for all $u \geq s\}$. There are $s_0(\eta)$ and $c_1 > 0$, depending only on K, R, λ, such that $\mathbb{P}(\mathcal{S}_\eta > s) \leq \exp(-c_1\eta^2\sqrt{s})$ for $s \geq s_0(\eta)$. In particular $\mathbb{E}[\mathcal{S}_\eta] < \infty$ and $\hat{\theta}_t \to \theta$ a.s.*

Proof. The map $t \mapsto C_\star\sqrt{\tilde{\beta}_t(\delta')}\, t^{-1/4}$ is eventually decreasing, since $\tilde{\beta}_t$ grows logarithmically while $t^{-1/4}$ decays polynomially; so on the event of Lemma E.7 the bound at $t = s$ controls all $t \geq s$. Using $\|\cdot\|_\infty \leq \|\cdot\|_2$, the conclusion $\|\hat{\theta}_u - \theta\|_\infty \leq \eta$ for all $u \geq s$ holds as soon as $C_\star^2\tilde{\beta}_s(\delta') \leq \eta^2\sqrt{s}$, i.e.

$$2\ln\tfrac{1}{\delta'} \;\leq\; \frac{4\eta^2\sqrt{s}}{C_\star^2} - (K-1)\ln\big(1 + \tfrac{2s}{\lambda}\big).$$

Let $s_0(\eta)$ be large enough that the subtracted logarithm is at most half the first term for $s \geq s_0(\eta)$; this is possible because $\sqrt{s}$ dominates $\ln s$. Then the choice $\ln(1/\delta') = \eta^2\sqrt{s}/C_\star^2 =: c_1\eta^2\sqrt{s}$ is admissible, and Lemma E.7 at that δ' gives $\mathbb{P}(\mathcal{S}_\eta > s) \leq \delta' = \exp(-c_1\eta^2\sqrt{s})$. Summability of a stretched exponential gives $\mathbb{E}[\mathcal{S}_\eta] \leq s_0(\eta) + \sum_{s\geq s_0} e^{-c_1\eta^2\sqrt{s}} < \infty$, and Borel–Cantelli gives the almost sure convergence. □

Table 3: The sandwich constant on the Arena instances, at the binding region of the top-1 claim. $\Delta(z_q)$ is realised, the Proposition F.14 column is the bound on it, and $D^\star$ is the Lemma F.12 envelope on the exact minimiser; the guarantee is $\psi_-(D^\star)$. The full and half pools agree to three decimals in every Δ column while λ_2 and S_A both halve, the residual is set by the margin, not the sample size, as Remark F.15 predicts.

instance	λ_2	S_A	$\Delta(z_q)$	Prop. F.14	$D^\star$	$\psi_-(D^\star)$
top-20, full pool	4134.5	76.6	0.226	0.254	0.437	**0.869**
top-20, replay half	2067.2	38.3	0.226	0.254	0.437	**0.869**
K=16, replay half	1789.5	88.2	0.361	0.421	0.760	**0.788**

Remark E.9. *The tail is stretched-exponential, $e^{-c\eta^2\sqrt{s}}$, rather than exponential in s. That is the correct order under a $\sqrt{t}$ exploration floor and is all the argument needs: only summability is used, in Lemma E.11.*

E.3 The constant, evaluated

$T^\dagger$ is computable offline. Proposition F.14 bounds the contrast width of the quadratic minimiser by $\Delta(z_q) \le g \max_{a<b} \sqrt{R_{ab}/R_{ij}}$ and Lemma F.12 converts it into the envelope $D^\star \ge \Delta(z_e)$. Evaluated at the binding region of the top-1 claim on the Arena fits:

Three things follow. The guaranteed inflation is $1.15\times$ on the top-20 instance and $1.27\times$ at $K = 16$, against the $4.08\times$ and $3.88\times$ the uniform form of Remark E.2 would give. The realised $\Delta(z_q)$ equals the margin g exactly in both cases, which is not coincidence: Proposition F.14 forces $(z_q)_j - (z_q)_i = g$, and on these instances the claim's own pair attains the widest contrast, so $\Delta(z_q) = g$ and the constant is essentially $\psi_-(g)$, *small exactly where certification is hard*. And the envelope is conservative by about an order of magnitude: the measured Λ_A/S_A of Section F.2 is 1.001–1.012, against the 1.15 the bound permits.

E.4 Step 3: regularity of the price

Fix $\varepsilon > 0$ and let $\Delta_\varepsilon = \{w \in \Delta_{\mathcal{P}} : \min_{ij} w_{ij} \ge \varepsilon\}$.

Lemma E.10. *Let A be a nonempty closed convex set defined by finitely many contrast half-spaces. On $\Delta_\varepsilon \times \Theta_R$:* (a) *$w \mapsto S_A(w;\vartheta)$ is concave and positively homogeneous of degree one;* (b) *$(w,\vartheta) \mapsto S_A(w;\vartheta)$ is finite and jointly continuous, and is differentiable in w wherever the projection's active set is unique, with $\partial S_A/\partial w_{ij} = \frac{1}{2}\sigma'(\vartheta_i - \vartheta_j)(z_i - z_j)^2$ at the optimal displacement z;* (c) *its super-gradients are bounded uniformly on $\Delta_\varepsilon \times \Theta_R$. Consequently $\Phi(\cdot;\vartheta)$ is concave, 1-homogeneous, jointly continuous and has bounded super-gradients, and Γ is continuous at θ.*

Proof. (a) $S_A(w;\vartheta) = \inf_{\theta' \in A} \frac{1}{2}\|\vartheta - \theta'\|^2_{L(w;\vartheta)}$ and L is linear in w, so each member of the infimum is linear and 1-homogeneous in w; an infimum of such is concave and 1-homogeneous.

(b) On $\Delta_\varepsilon \times \Theta_R$, $L(w;\vartheta) \succeq a\varepsilon K I_V$ on V, so the inner problem is a strictly convex quadratic program over a fixed nonempty polyhedron in $\mathbb{R}^K/\mathbb{R}\mathbf{1}$, with a unique minimiser; finiteness is immediate. Joint continuity follows from Berge's maximum theorem, the constraint correspondence being constant and the objective jointly continuous with compact effective sublevel sets by the uniform positive-definiteness. Differentiability at points of unique active set is Danskin's theorem applied to an infimum of functions linear in w; differentiating $\frac{1}{2}\|z\|^2_{L(w;\vartheta)}$ in w_{ij} at fixed z gives the display.

(c) Any super-gradient is a convex combination of vectors with entries $\frac{1}{2}\sigma'(\vartheta_i - \vartheta_j)(z_i - z_j)^2$. Since $\sigma' \le \frac{1}{4}$ and $\|z\|_2^2 \le 2S_A/(a\varepsilon K)$ with S_A bounded on the compact set, the entries are bounded by a constant depending only on K, R, ε.

The consequences for Φ hold because a minimum of finitely many concave, 1-homogeneous, continuous functions with bounded super-gradients inherits those properties, its super-differential at w being the convex hull of those of the members attaining the minimum; $\Gamma = \max_w \Phi$ is continuous by Berge again, $\Delta_{\mathcal{P}}$ being fixed and compact. Note $\mathcal{W}$ is the fixed family $\mathcal{W}(\theta)$ throughout. □

E.5 Step 4: the design reaches the maximiser

Lemma E.11. *Let $\varepsilon > 0$. There is a random time $\mathcal{T}_\varepsilon$ with $\mathbb{E}[\mathcal{T}_\varepsilon] < \infty$ such that for $t \geq \mathcal{T}_\varepsilon$, $\Phi(\hat{w}_t; \theta) \geq \Gamma(\theta) - \varepsilon$ and $\|\hat{\theta}_t - \theta\|_\infty \leq r_0$. In particular $\hat{w}_t \to \arg\max_w \Phi(w; \theta)$ a.s.*

Proof. By Lemma E.10, $\Phi(\cdot; \theta)$ is concave, 1-homogeneous with bounded super-gradients on each $\Delta_{\varepsilon'}$; by Lemma E.4 the design satisfies $\hat{w}_t \in \Delta_{\varepsilon'_t}$ with $\varepsilon'_t = (\sqrt{t}/(2m) - 1)/t > 0$ for large t, so the iterates stay in the relative interior. Steps (i)–(iii) are a Frank–Wolfe iteration on $\Phi(\cdot; \hat{\theta}_t)$ with step $1/(t+1)$.

Two facts make the window of step (ii) a legitimate super-gradient oracle. The family is fixed once so that $\|\hat{\theta}_t - \theta\|_\infty \leq r_0$, so the minimum is over a fixed finite index set. And the super-differential of a minimum of concave functions is the convex hull of the gradients of the *active* members; the window contains every member within $(1 + t^{-0.9})$ of the minimum, hence all active ones, and $t^{-0.9} \to 0$, so the average G_t is an element of the super-differential up to an error tending to zero. This is what the window is for: where several regions bind, the single-argmin gradient is not a super-gradient and the vertex step chatters.

By the uniform continuity supplied by Lemma E.10(b) on the compact set $\Delta_{\varepsilon'} \times \{\|\vartheta - \theta\|_\infty \leq r_0\}$, choose $\eta \in (0, r_0]$ with

$$\|\vartheta - \theta\|_\infty \leq \eta \implies \sup_{w \in \Delta_{\varepsilon'}} \|\nabla_w \Phi(w; \vartheta) - \nabla_w \Phi(w; \theta)\|_\infty \leq \omega(\varepsilon), \tag{1}$$

with $\omega(\varepsilon)$ small enough that the Frank–Wolfe theorem for concave 1-homogeneous objectives with bounded super-gradients and an ω-inexact oracle (Wang et al., 2021) returns $\Phi(\hat{w}_t; \theta) \geq \Gamma(\theta) - \varepsilon$ for all $t \geq T_{\mathrm{FW}}(\varepsilon)$, a *deterministic* time depending only on ε, K, R and the bound of Lemma E.10(c).

Set $\mathcal{T}_\varepsilon = \mathcal{S}_\eta + T_{\mathrm{FW}}(\varepsilon)$ with $\mathcal{S}_\eta$ as in Corollary E.8. From round $\mathcal{S}_\eta$ onwards the oracle is $\omega(\varepsilon)$-inexact by equation 1, so the displayed conclusion holds from $\mathcal{S}_\eta + T_{\mathrm{FW}}(\varepsilon)$. Finally $\mathbb{E}[\mathcal{T}_\varepsilon] = T_{\mathrm{FW}}(\varepsilon) + \mathbb{E}[\mathcal{S}_\eta] < \infty$ by Corollary E.8, this is the only place the tail of $\mathcal{S}_\eta$ is used, and summability suffices. Letting $\varepsilon \downarrow 0$ along a sequence and applying Borel–Cantelli gives $\Phi(\hat{w}_t; \theta) \to \Gamma(\theta)$ a.s.; since $\Phi(\cdot; \theta)$ is concave with compact argmax on the compact $\Delta_{\mathcal{P}}$, every limit point of $\hat{w}_t$ is a maximiser. □

E.6 Step 5: a deterministic stopping time, and assembly

Lemma E.12. *Let $\varepsilon \in (0, \Gamma(\theta))$, let $g(t)$ denote the non-$\ln(1/\delta)$ part of the boundary, so that $\beta(t, \delta) = \ln(1/\delta) + g(t)$ with $g(t) = o(t)$. Define*

$$T_\varepsilon(\delta) = \inf\big\{t \geq 1 :\ \psi\, t\,(\Gamma(\theta) - \varepsilon)\ \geq\ \ln(1/\delta) + g(t)\big\}.$$

Then $T_\varepsilon(\delta) < \infty$ and $\lim_{\delta \to 0} T_\varepsilon(\delta)/\ln(1/\delta) = T^\dagger + \varepsilon$.

Proof. Write $h(t) = t/(T^\dagger + \varepsilon) - g(t)$. Since $g(t) = o(t)$ and $T^\dagger + \varepsilon < \infty$, $h(t)/t \to (T^\dagger + \varepsilon)^{-1} > 0$, so $h(t) \to \infty$ and $T := T_\varepsilon(\delta)$ is finite. By minimality $h(T) \geq \ln(1/\delta) > h(T-1)$, so

$$\frac{h(T)}{T}\ \geq\ \frac{\ln(1/\delta)}{T}\ >\ \frac{h(T-1)}{T-1} \cdot \frac{T-1}{T}.$$

As $\delta \to 0$, $T \to \infty$ and both outer terms tend to $(T^\dagger + \varepsilon)^{-1}$, giving the claim. □

Proof of Theorem 3.3. Fix $\varepsilon \in (0, \Gamma(\theta)/2)$. By Lemma E.10, shrink $\eta \in (0, r_0]$ so that additionally

$$\|\vartheta - \theta\|_\infty \leq \eta \implies \sup_{w \in \Delta_{\mathcal{P}}} |\Phi(w; \vartheta) - \Phi(w; \theta)| \leq \varepsilon/2, \tag{2}$$

possible because Φ is jointly continuous, hence uniformly so, on the compact $\Delta_{\mathcal{P}} \times \{\|\vartheta - \theta\|_\infty \leq r_0\}$. Let $\mathcal{T}_{\varepsilon/2}$ be the time of Lemma E.11 for tolerance $\varepsilon/2$ and this η.

For $t \geq \mathcal{T}_{\varepsilon/2}$ we have $\mathcal{W}(\hat{\theta}_t) = \mathcal{W}$ and by 1-homogeneity, Lemma E.11 and equation 2,

$$\min_{A \in \mathcal{W}} S_A(N_t; \hat{\theta}_t) = t\,\Phi(\hat{w}_t; \hat{\theta}_t)\ \geq\ t\big(\Phi(\hat{w}_t; \theta) - \tfrac{\varepsilon}{2}\big)\ \geq\ t\big(\Gamma(\theta) - \varepsilon\big).$$

The rule stops on Λ_A, so apply Theorem F.10 *to each region separately*, before taking any minimum: for all large t, $\Lambda_A(N_t) \geq \psi_-(\bar{\Delta}_A)\, S_A(N_t;\hat{\theta}_t) = \psi_-(\bar{\Delta}_A)\, t\, S_A(\hat{w}_t;\hat{\theta}_t)$ by 1-homogeneity. The rule fires once every region has crossed, so the binding requirement is $t \geq \beta(t,\delta)/\big[\psi_-(\bar{\Delta}_A) S_A(\hat{w}_t;\hat{\theta}_t)\big]$ for the worst A, i.e. $t \geq T^\dagger \beta(t,\delta)$ up to the ε slack already accumulated. Bounding each region by its own ψ_- rather than by a common one is what separates $T^\dagger$ from $T^*/\psi_-(\max_A \bar{\Delta}_A)$ (Remark E.2). *No concentration is used in this step: both sides are functions of the observed* $(N_t, \hat{\theta}_t)$. If in addition $t \geq T_\varepsilon(\delta)$ then $\psi t(\Gamma(\theta) - \varepsilon) \geq \ln(1/\delta) + g(t) = \beta(t,\delta)$, so every region has crossed and the rule fires. Hence pointwise

$$\tau_\delta \;\leq\; \max\{\mathcal{T}_{\varepsilon/2}, T_\varepsilon(\delta)\} \;\leq\; \mathcal{T}_{\varepsilon/2} + T_\varepsilon(\delta).$$

Take expectations, use $\mathbb{E}[\mathcal{T}_{\varepsilon/2}] < \infty$, divide by $\ln(1/\delta)$ and let $\delta \to 0$: the first term vanishes as it does not depend on δ, and Lemma E.12 gives $\limsup_\delta \mathbb{E}[\tau_\delta]/\ln(1/\delta) \leq T^\dagger + \varepsilon$. The left side does not depend on ε; let $\varepsilon \downarrow 0$. (Stopping on S_A at the uninflated β would remove the appeal to Theorem F.10, but that rule is not δ-correct; see Remark E.3.) □

Remark E.13 (What a matching lower bound would require)**.** *Theorem 3.3 is an upper bound only, and we do not claim asymptotic optimality. The standard change-of-measure argument delivers* $\liminf_\delta \tau_\delta/\ln(1/\delta) \geq T^*_{\mathrm{KL}}$ *with* $1/T^*_{\mathrm{KL}} = \max_w \min_A \inf_{\theta' \in A} \sum_{ij} w_{ij} \mathrm{KL}_{ij}(\theta\|\theta')$*, whose inner quantity is the limit of* $\Lambda_A(N_t)/t$*, not of* $S_A(N_t)/t$*. By Theorem F.10 these differ by a factor in* $[\psi_-(\bar{\Delta}), \psi_+(\bar{\Delta})]$*, and Remark F.15 shows the bracket does not close as* $t \to \infty$*: it is set by the margin, not the sample size. Hence* $T^* = \Gamma^{-1}$*, built from the quadratic price, is in general* not T^*_{KL}*, and the two bounds do not meet at a common constant. Two routes would close this, and we have taken neither: route on the exact price, replacing the Danskin gradient of* S_A *by that of* Λ_A*, so that both the design and the lower bound refer to the same functional; or retain the quadratic routing and prove the weaker statement that the achieved rate lies within* $[\psi_-, \psi_+]$ *of optimal, which the sandwich already supplies. We regard the second as the honest reading of what is proved here.*

F THE STOPPING RULE AND ITS CORRECTNESS

Items $[K]$, unordered pairs $\mathcal{P}$, $\sigma(x) = (1+e^{-x})^{-1}$, $A_{\log}(\eta) = \log(1+e^\eta)$ so that $A'_{\log} = \sigma$ and $A''_{\log} = \sigma'$. With W_{ab} the count of a-beats-b and $n_{ab} = W_{ab} + W_{ba}$, the Bradley–Terry log-likelihood at counts N is, writing $\eta_{ab}(\theta) = \theta_a - \theta_b$,

$$\ell_N(\theta) \;=\; \sum_{a<b} \Big[W_{ab}\, \eta_{ab}(\theta) - n_{ab} A_{\log}\big(\eta_{ab}(\theta)\big) \Big].$$

Let $\hat{\theta}$ maximise ℓ_N on $E = \{\mathbf{1}^\top \theta = 0\}$ and write $\hat{\eta}_{ab} = \hat{\theta}_a - \hat{\theta}_b$. The observed information is the weighted Laplacian $L(N) = \sum_{a<b} n_{ab} \sigma'(\hat{\eta}_{ab})(e_a - e_b)(e_a - e_b)^\top = -\nabla^2 \ell_N(\hat{\theta})$, and $V_t = \sum_{s \leq t} x_s x_s^\top$ is its unweighted counterpart. For $z \in \mathbb{R}^K$ the **contrast width** is $\Delta(z) := \max_{a<b} |z_a - z_b|$; it is the only quantity the residual below depends on.

Definition F.1 (The two statistics)**.** *For closed convex* $A \subseteq E$*,*

$$\underbrace{\Lambda_A(N_t) = \ell_t(\hat{\theta}_t) - \sup_{\theta' \in A} \ell_t(\theta')}_{\textit{exact log-GLR}}, \qquad \underbrace{S_A(N_t) = \tfrac{1}{2} \inf_{\theta' \in A} \|\theta' - \hat{\theta}_t\|^2_{L(N_t)}}_{\textit{quadratic price}}.$$

Λ_A is what the stopping rule tests; S_A is what the projection subroutine returns and what the sampling rule differentiates.

This appendix proves:

(i) Stopping on Λ_A at the threshold β of Definition F.2 is δ-correct for every sampling rule and every witness system, with no union over regions, claims or times (Theorem 3.2, Corollary F.7). β carries a $\log\det$ over all $K-1$ contrast directions. This is the rule every experiment in this paper runs on.

(ii) Λ_A and S_A agree only up to a bracket set by the *margin* and not by the sample size (Theorem F.10, Remark F.15); stopping on S_A inflated by a computable factor inherits the guarantee (Corollary F.16), and stopping on S_A *without* that factor does not.

F.1 THE GUARANTEE

Definition F.2 (Stopping rule; restating Definition 3.1)**.** *Fix $\delta \in (0,1)$ and $\lambda > 0$ in advance, and let*

$$\beta(t,\delta) \;=\; \ln\frac{1}{\delta} + \frac{1}{2}\ln\det_E\Big(I + \frac{V_t}{4\lambda}\Big) + \frac{\lambda}{2}\|\hat\theta_t\|^2.$$

Let $\tau = \inf\{t \ge 1 :\ \hat\theta_t$ exists and $\Lambda_A(t) \ge \beta(t,\delta)$ for every $A \in \mathcal{W}(\hat\theta_t)\}$, and return $\mathrm{ans}(\hat\theta_\tau)$ if $\tau < \infty$. This is an $(\mathcal{F}_t)$-stopping time, since $\{\tau = t\}$ is a function of H_t.

Lemma F.3 (Likelihood-ratio martingale)**.** *Let $\pi = N(0, \lambda^{-1} I_E)$ on E and $M_t = \int_E \exp(\ell_t(\theta') - \ell_t(\theta))\,\pi(d\theta')$. Under the Bradley–Terry model, (M_t) is a nonnegative $(\mathcal{F}_t)$-martingale under $\mathbb{P}_\theta$ with $M_0 = 1$, for every sampling rule; hence $\mathbb{P}_\theta(\exists t :\ M_t \ge 1/\delta) \le \delta$.*

Proof. Fix θ' and put $r_s = p_{\theta'}(Y_s \mid x_s)/p_\theta(Y_s \mid x_s)$, so $\exp(\ell_t(\theta') - \ell_t(\theta)) = \prod_{s\le t} r_s$. Since x_s is $\mathcal{F}_{s-1}$-measurable and $p_\theta > 0$, $\mathbb{E}_\theta[r_s \mid \mathcal{F}_{s-1}] = \sum_y p_\theta(y \mid x_s) p_{\theta'}(y \mid x_s)/p_\theta(y \mid x_s) = 1$: the probabilities with which the rule chooses pairs are the same under θ' and θ and cancel, which is why no property of the sampling rule is needed. Joint measurability and nonnegativity give, by conditional Tonelli, $\mathbb{E}_\theta[M_t \mid \mathcal{F}_{t-1}] = M_{t-1}$; and $M_0 = 1$ since $\ell_0 \equiv 0$. Finiteness: $\ell_t(\theta') \le 0$, so $M_t \le e^{-\ell_t(\theta)} < \infty$. Ville's inequality gives the last claim. □

Lemma F.4 (Global quadratic bound)**.** *If $\hat\theta_t$ exists, then $\ell_t(\theta') \ge \ell_t(\hat\theta_t) - \frac{1}{8}\|\theta' - \hat\theta_t\|^2_{V_t}$ for every $\theta' \in E$.*

Proof. Let $f(u) = \ell_t(\hat\theta_t + u(\theta' - \hat\theta_t))$. Then $f'(0) = 0$ by stationarity and $f''(u) \ge -\frac{1}{4}\|\theta' - \hat\theta_t\|^2_{V_t}$ since $\sigma' \le \frac{1}{4}$. Integrating twice from 0 to 1 gives the claim. □

Lemma F.5 (Gaussian integral)**.** *For $Q \succeq 0$ on E, $\mu \in E$, $\lambda > 0$,*

$$\int_E e^{-\frac{1}{2}\|\theta' - \mu\|^2_Q}\,\pi(d\theta') = \det_E\Big(I + \tfrac{Q}{\lambda}\Big)^{-1/2} e^{-\frac{\lambda}{2}\mu^\top Q(Q+\lambda I)^{-1}\mu} \;\ge\; \det_E\Big(I + \tfrac{Q}{\lambda}\Big)^{-1/2} e^{-\frac{\lambda}{2}\|\mu\|^2}.$$

Proof. The exponent against π's density is minimised at $\theta^* = (Q+\lambda I)^{-1}Q\mu$ with value $\frac{\lambda}{2}\mu^\top Q(Q+\lambda I)^{-1}\mu$; the residual Gaussian integral contributes $\det_E(I + Q/\lambda)^{-1/2}$. The inequality is $Q(Q+\lambda I)^{-1} \preceq I$. No invertibility of Q is used. □

Theorem F.6 (δ-correctness; restating Theorem 3.2)**.** *Under the Bradley–Terry model, for any sampling rule and any witness system, $\mathbb{P}_\theta(\tau < \infty$ and $\mathrm{ans}(\hat\theta_\tau) \ne \mathrm{ans}(\theta)) \le \delta$.*

Proof. Step 1. On the bad event $\theta \in \mathrm{Alt}(\hat\theta_\tau)$, so by completeness $\theta \in A$ for some $A \in \mathcal{W}(\hat\theta_\tau)$. The rule stopped, so $\Lambda_A(\tau) \ge \beta(\tau,\delta)$; and $\theta \in A$ gives $\sup_{\theta' \in A} \ell_\tau(\theta') \ge \ell_\tau(\theta)$, hence $\Lambda_A(\tau) \le \ell_\tau(\hat\theta_\tau) - \ell_\tau(\theta)$. So the bad event is contained in

$$\mathcal{E} = \{\exists t :\ \hat\theta_t \text{ exists and } \ell_t(\hat\theta_t) - \ell_t(\theta) \ge \beta(t,\delta)\},$$

which mentions no region. This step is where the exactness of Λ_A is used, and it has no analogue for S_A (Remark F.11).

Step 2. Lemma F.4 then Lemma F.5 with $Q = V_t/4$ and $\mu = \hat\theta_t$ give $M_t \ge e^{\ell_t(\hat\theta_t) - \ell_t(\theta)} \det_E(I + V_t/4\lambda)^{-1/2} e^{-\frac{\lambda}{2}\|\hat\theta_t\|^2}$, i.e. $\ell_t(\hat\theta_t) - \ell_t(\theta) \le \ln M_t + \beta(t,\delta) - \ln(1/\delta)$ pointwise. The inequality holds for each ω and each t, so the randomness of β is immaterial. On $\mathcal{E}$ it forces $M_t \ge 1/\delta$ for some t, and $\mathbb{P}_\theta(\mathcal{E}) \le \delta$ by Lemma F.3. □

Remark F.7 (Anytime validity, all claims at once)**.** *Run any number of answer maps ans_c with witness systems $\mathcal{W}_c$ on the same votes, and say $\mathrm{ans}_c(\hat\theta_t)$ passes at t if $\Lambda_A(t) \ge \beta(t,\delta)$ for every $A \in \mathcal{W}_c(\hat\theta_t)$. With probability at least $1 - \delta$ no false assertion passes at any time. No union over claims, regions or times is needed.*

F.2 WHAT THE QUADRATIC PRICE COSTS

Theorem 3.2 governs Λ_A. The projection subroutine returns S_A. The two are related by an identity plus a scalar comparison, and by nothing else.

Lemma F.8 (Bregman identity). *For $\theta' \in \mathbb{R}^K$, $z = \theta' - \hat{\theta}$, $\delta_{ab} = z_a - z_b$,*

$$\ell_N(\hat{\theta}) - \ell_N(\theta') = \sum_{a<b} n_{ab}\, D(\hat{\eta}_{ab} + \delta_{ab} \| \hat{\eta}_{ab}), \quad D(\eta\|\hat{\eta}) := A_{\log}(\eta) - A_{\log}(\hat{\eta}) - \sigma(\hat{\eta})(\eta - \hat{\eta}),$$

with every term nonnegative. All the approximation is thereby confined to one scalar comparison per pair.

Proof. Expanding and adding and subtracting $n_{ab}\sigma(\hat{\eta}_{ab})\delta_{ab}$ gives the displayed sum plus $T = \sum_{a<b}(n_{ab}\sigma(\hat{\eta}_{ab}) - W_{ab})\delta_{ab} = -\langle z, \nabla\ell_N(\hat{\theta})\rangle$. The MLE is stationary on E and $\nabla\ell_N$ is orthogonal to $\mathbf{1}$ since every term is a contrast, so $\nabla\ell_N(\hat{\theta}) = 0$ and $T = 0$. Nonnegativity of D is convexity of $A_{\log}$. □

Write $\psi_-(x) = 2(e^{-x} + x - 1)/x^2$ and $\psi_+(x) = 2(e^x - x - 1)/x^2$, with $\psi_\pm(0) = 1$; ψ_- decreases from 1, ψ_+ increases from 1, and $\psi_\pm(x) = 1 \pm x/3 + O(x^2)$.

Lemma F.9 (Self-concordance). $\frac{1}{2}\psi_-(|\delta|)\sigma'(\hat{\eta})\delta^2 \le D(\hat{\eta} + \delta\|\hat{\eta}) \le \frac{1}{2}\psi_+(|\delta|)\sigma'(\hat{\eta})\delta^2$.

Proof. $\frac{d}{d\eta}\log\sigma'(\eta) = -\tanh(\eta/2)$ has modulus at most one, so $\log\sigma'$ is 1-Lipschitz and $e^{-|u|}\sigma'(\hat{\eta}) \le \sigma'(\hat{\eta} + u) \le e^{|u|}\sigma'(\hat{\eta})$. Taylor with integral remainder gives $D = \int_0^\delta (\delta - u)\sigma'(\hat{\eta} + u)\, du$; substituting the two bounds and using $\int_0^{|\delta|}(|\delta| - u)e^{\mp u}\, du = e^{\mp|\delta|} \pm |\delta| - 1$ yields the claim. □

Theorem F.10 (Sandwich). *Let z_q minimise S_A and z_e minimise Λ_A, both unique when $L(N) \succ 0$ on E and A is nonempty closed convex. Then $\psi_-(\Delta(z_e))\, S_A \le \Lambda_A \le \psi_+(\Delta(z_q))\, S_A$.*

Proof. Upper. $\hat{\theta} + z_q$ is feasible for Λ_A, so by Lemmas F.8 and F.9 and $|(z_q)_{ab}| \le \Delta(z_q)$, $\Lambda_A \le \frac{1}{2}\psi_+(\Delta(z_q))\sum n_{ab}\sigma'(\hat{\eta}_{ab})(z_q)_{ab}^2 = \psi_+(\Delta(z_q))S_A$. *Lower.* The same two lemmas at z_e give $\Lambda_A \ge \frac{1}{2}\psi_-(\Delta(z_e))\|z_e\|^2_{L(N)}$, and $\frac{1}{2}\|z_e\|^2_{L(N)} \ge S_A$ since $\hat{\theta} + z_e \in A$. □

Remark F.11 (Stopping on S_A is not valid without the correction). *Since $\psi_- < 1$, the sandwich places S_A in $[\Lambda_A/\psi_+, \Lambda_A/\psi_-]$, whose upper end* exceeds *Λ_A by up to 7% at $\Delta = 0.22$ and 12% at $\Delta = 0.35$, the margins these experiments run at. Testing $S_A \ge \beta$ can therefore fire while Λ_A is only $\psi_-\beta < \beta$, before Step 1 of Theorem 3.2 licenses it. Step 1 rests on the identity $\Lambda_A \le \ell_t(\hat{\theta}_t) - \ell_t(\theta)$, which holds for the exact log-GLR and has no analogue for the surrogate. The valid route is Corollary F.16.*

Only the lower bound refers to the unknown z_e; the next lemma removes it.

Lemma F.12 (Envelope). *Let $\lambda^+_{\min} > 0$ be the smallest nonzero eigenvalue of $L(N)$ on E and $D^\star = \min\{D > 0 : \frac{\lambda^+_{\min}}{4}(e^{-D} + D - 1) > \psi_+(\Delta(z_q))S_A\}$. Then $\Delta(z_e) < D^\star$ and $\Lambda_A \ge \psi_-(D^\star)S_A$.*

Proof. For $\theta' \in A$ with $z = \theta' - \hat{\theta}$ taken in E, $\Delta(z) \le 2\|z\|_2$ so $\|z\|_2^2 \ge \Delta^2/4$, and Lemmas F.8–F.9 with ψ_- decreasing give $\ell_N(\hat{\theta}) - \ell_N(\theta') \ge \frac{\lambda^+_{\min}}{8}\psi_-(\Delta)\Delta^2 = \frac{\lambda^+_{\min}}{4}(e^{-\Delta} + \Delta - 1)$, increasing and unbounded in Δ. So $\Delta(z_e) \ge D^\star$ would contradict Theorem F.10. The last claim is ψ_- decreasing. □

Remark F.13 (Why a spectral constant, and not a single pair). *Bounding $\|z\|^2_{L(N)}$ below by the single pair attaining Δ gives $\|z\|^2_{L(N)} \ge n_{ab}\sigma'(\hat{\eta}_{ab})\Delta^2$ and so replaces the constant $\lambda^+_{\min}/4$ by* $\min_{ab} n_{ab}\sigma'(\hat{\eta}_{ab})$. *On the Arena top-*$20$ *fit that minimum is* zero*: three of the* 190 *pairs were never played, and an unplayed pair leaves Δ entirely uncharged, so the bound is vacuous and $D^\star$ runs to the search cap. Restricted to the* 187 *observed pairs the constant is* 46.9 *against $\lambda^+_{\min}/4 = 1033.6$,*

and $D^\star$ degrades from 0.429 to 2.698 — ψ_- from 0.871 to 0.485, so a 15% correction becomes a 106% one. The reason is structural: the single-pair bound charges the whole displacement to the least-informative pair, whereas a displacement of large contrast width necessarily spreads L-norm across the graph, which is what $\lambda^+_{\min}$ measures.

Proposition F.14 (Δ for a single crossing)**.** *For $A = \{\theta'_j \geq \theta'_i\}$ with margin $g = \hat\theta_i - \hat\theta_j > 0$ and $R_{ab} = (e_a - e_b)^\top L(N)^+(e_a - e_b)$, the quadratic minimiser is $z_q = (g/R_{ij})L(N)^+(e_j - e_i)$, with $(z_q)_j - (z_q)_i = g$, $|(z_q)_a - (z_q)_b| \leq g\sqrt{R_{ab}/R_{ij}}$, and $S_A = g^2/(2R_{ij})$. Hence $\Delta(z_q) \leq g \max_{a<b}\sqrt{R_{ab}/R_{ij}}$.*

Proof. KKT for projection onto a half-space with normal $y = e_j - e_i$ gives $z = \lambda L^+ y$ with $-g + \lambda y^\top L^+ y = 0$, so $\lambda = g/R_{ij}$, $\frac{1}{2}z^\top Lz = g^2/(2R_{ij})$ and $\langle y, z\rangle = g$. For $u = e_a - e_b$, $|(z_q)_a - (z_q)_b| = \lambda|u^\top L^+ y| \leq g\sqrt{R_{ab}R_{ij}}/R_{ij}$ by Cauchy–Schwarz in the L^+ inner product. □

Remark F.15 (The residual does not vanish with more votes)**.** *Under a fixed design with $\min_{ab} w_{ab} > 0$ every resistance scales as $1/t$, so the* ratios R_{ab}/R_{ij} *converge and $\Delta(z_q)$ tends to a nonzero constant. S_A and Λ_A both grow linearly in t while their ratio tends to a constant in $[\psi_-, \psi_+]$ that is in general not 1. What controls that factor is the* margin*, not the number of votes. This is visible directly in the data: on the Arena top-20, halving the pool halves $\lambda^+_{\min}$ ($4134.5 \to 2067.2$) and halves S_A ($76.6 \to 38.3$) while leaving $\Delta(z_q) = 0.226$ and $D^\star = 0.429$ unchanged to three decimals (Table 3). Since $\psi_\pm(x) = 1 \pm x/3 + O(x^2)$, the bracket is within $\pm 13\%$ for $\Delta \leq 0.35$ — and by Proposition F.14 $\Delta(z_q)$ is small exactly when g is, so* the residual is smallest where certification is hardest*. It is loose for well-separated models ($\psi_- = 0.57$ at $\Delta = 2$) and vacuous if the vote graph is disconnected, since then $R_{ij} = \infty$ and $S_A = 0$; the forced-exploration step prevents that, supplying consistency for the rate* and *keeping this bound usable.*

Remark F.16 (Certified stopping on S_A)**.** *Replacing the test $S_A \geq \beta(t,\delta)$ by $S_A \geq \beta(t,\delta)/\psi_-(D^\star)$ gives $\Lambda_A \geq \beta(t,\delta)$ whenever it fires, so Theorem 3.2 applies. The factor $1/\psi_-(D^\star)$ is computed at run time from quantities the projection already returns, and costs 15% of the budget on the Arena top-20 fit ($D^\star = 0.429$, $\psi_- = 0.871$) and 27% at $K = 16$ ($D^\star = 0.752$, $\psi_- = 0.790$). Most of that is slack in the envelope of Lemma F.12 rather than a real effect: the realised contrast widths are 0.226 and 0.361, at which ψ_- is 0.929 and 0.890, so a 7–11% correction is inflated to 15–27% by bounding the unknown $\Delta(z_e)$.*

Measured. Top-20 Arena fit, the two binding regions. Λ_A is the exact constrained log-GLR, computed by Newton on the likelihood with θ_a and θ_b merged (the minimiser of a convex objective over a half-space whose interior excludes $\hat\theta$ lies on the boundary); S_A comes from the projection subroutine. The upper block scales the design by a constant, which leaves $\hat\theta$ and every resistance ratio fixed and so isolates the dependence on t alone; the lower block thins the pool binomially and refits $\hat\theta$, so the margin moves too (`psi_measure.py`).

t	pair	g	S_A	Λ_A	Λ_A/S_A	$\Delta(z_q)$	$[\psi_-, \psi_+]$
design scaled by a constant; $\hat\theta$ fixed							
6,708	10v13	0.226	1.445	1.446	1.0011	0.226	[0.929, 1.080]
6,708	10v8	0.277	2.710	2.732	1.0081	0.277	[0.914, 1.099]
35,551	10v13	0.226	7.656	7.665	1.0011	0.226	[0.929, 1.080]
35,551	10v8	0.277	14.364	14.479	1.0081	0.277	[0.914, 1.099]
177,754	10v13	0.226	38.281	38.324	1.0011	0.226	[0.929, 1.080]
177,754	10v8	0.277	71.818	72.397	1.0081	0.277	[0.914, 1.099]
355,507	10v13	0.226	76.562	76.648	1.0011	0.226	[0.929, 1.080]
355,507	10v8	0.277	143.636	144.794	1.0081	0.277	[0.914, 1.099]
binomially thinned; $\hat\theta$ refit							
6,845	10v13	0.217	1.432	1.435	1.0020	0.217	[0.931, 1.077]
6,845	10v5	0.237	1.719	1.725	1.0032	0.237	[0.925, 1.084]
35,433	10v13	0.252	9.470	9.484	1.0014	0.252	[0.921, 1.090]
35,433	10v5	0.350	19.725	19.848	1.0062	0.350	[0.893, 1.128]
177,899	10v13	0.216	35.659	35.697	1.0011	0.216	[0.932, 1.076]
177,899	10v8	0.240	54.304	54.654	1.0064	0.240	[0.925, 1.085]

In the upper block the ratio is constant to four decimals across a 53-fold range in t: with $\hat{\theta}$ held fixed, S_A and Λ_A both scale exactly linearly and $\Delta(z_q)$ does not move at all, which is Remark F.15 in its sharpest form. In the lower block the ratio does move but with the margin, not with t: its largest value, 1.0062, occurs at the *smallest* of the three budgets that carries the widest contrast ($\Delta = 0.350$), and at $t = 177{,}899$ the two regions differ by a factor of six in $\Lambda_A - S_A$ while sharing a budget.

Two things are worth separating. The ratio sits well inside the bracket: the realised discrepancy is 0.1–0.8% against a bracket of ± 7–13%, so the sandwich is loose by an order of magnitude on this instance. And the ratio exceeds 1 in every row here S_A *understates* the likelihood drop, so stopping on S_A would have been conservative. That second observation is an empirical fact about one fit, not a theorem. The bracket permits S_A to overstate Λ_A by 8% at $\Delta = 0.226$ and 12% at $\Delta = 0.350$, which would be anti-conservative (Remark F.11), and nothing in the fit rules that out at another instance. Accordingly we stop on Λ_A and route on S_A, where approximation error cannot affect correctness; and we do not report $g^2/(2R_{ij})$ as a likelihood drop without the bracket attached.

G Empirical results

This appendix collects every measurement in the paper. It is organised in the order the questions arise: what the procedures are (§G.1); what the votes a leaderboard has already collected establish (§G.2) and how that compares with standard tests of ranking (§G.3); what certification costs against current practice (§G.4); whether exact joint pricing pays (§G.5); the same headline on real recorded votes rather than resampled ones (§G.6); comparisons against state-of-the-art routing (§G.7), against the concurrent optimal top-k algorithm (§G.8) and against non-parametric baselines (§G.9); and a live deployment in which votes are produced on demand (§G.10).

Every cell reported here stops on the proved boundary of Theorem 3.2. An earlier version of this appendix reported several cells under a conjectured one-dimensional boundary; those numbers are withdrawn, and where the conclusion changed we say so at the point it changed rather than only here.

G.1 Protocol and methods compared

Methods compared. Every table below compares the same four procedures, which differ only in how they compute a claim's transportation cost and how they choose the next vote.

Certify the procedure of this paper: exact transportation costs for every alternative region, the stopping rule of Theorem 3.2, and votes routed by the alternative-gradient rule of Section 3.

Relaxed identical, except that a joint claim's transportation cost is replaced by that of its cheapest single condition rather than by the projection onto their intersection — the *weakest-link relaxation*. Since a dominated model must escape all of its dominators at once, this underestimates the claim's cost and so stops later; the column isolates what exact joint pricing buys.

Uniform+GLR the same stopping rule, with votes drawn uniformly over pairs instead of by alternative-set gradients. The column isolates what *routing* buys, separately from the stopping rule.

Elo-CI current practice: uniform voting, per-model Wald confidence intervals with a union bound over pairs, certifying a claim when the two intervals separate.

GLR abbreviates *generalised likelihood ratio*: the statistic of Theorem 3.2, the log-likelihood distance from the fitted score vector to the nearest point of the claim's alternative region.

Ground truth from real battles. We fit Bradley–Terry scores from the public Chatbot Arena dataset: $1{,}157{,}137$ anonymous decisive battles, May-2024 era, spanning 129 models. The top-20 models by battle count account for $355{,}507$ battles among themselves, and those are the votes the replays consume. The fit reproduces the published Elo table (e.g. GPT-4o at 1631). In §§G.4–G.5 votes are simulated from the fitted model so that correctness can be scored against a known truth while the vote-generating process matches real preference data; §G.6 onwards replaces this with actual recorded votes. API list rates ($/M output tokens) supply the cost coordinate; the true quality–cost frontier comprises five models. All predictions were fixed before the first full run. Unless stated otherwise cells use 25 seeds; we report wins and losses symmetrically, and mark as

Table 4: Fraction of fitted pairwise orderings certified by the recorded Arena votes. "Simultaneous" certifies all reported comparisons jointly; "per-comparison" assigns each its own 95% level; "proved" uses the threshold of Theorem 3.2, which is also the threshold every experiment in this paper stops on. Denominators count *every* position the ordering asserts: at top-64, 11 of the 63 consecutive positions have no direct votes and are priced through the graph like any other pair.

	all ranked pairs	*consecutive* pairs		
models	simultaneous	simultaneous	per-comparison 95%	proved
top-20	173/190 (91.1%)	7/19 (36.8%)	10/19 (52.6%)	1/19 (5.3%)
top-40	713/780 (91.4%)	6/39 (15.4%)	16/39 (41.0%)	0/39 (0.0%)
top-64	1827/2016 (90.6%)	**3/63 (4.8%)**	15/63 (23.8%)	**0/63 (0.0%)**

censored any cell that reaches its budget without certifying, since such a cell bounds performance from below and cannot rank methods.

Two fields. Results are reported on two model sets. The *top*-20 field is the one the audit of §G.2 concerns. Under the proved boundary the top-1 claim on it is not reachable from the archive, so every routing rule censors there and the comparison is empty; §G.6 shows this directly. The $K{=}16$ field, the fitted leader plus the 15 lowest-ranked of the top-20, i.e. the four closest challengers dropped, is the largest subset on which the claim is within reach (134% of β on the replay half). It was fixed before the runs, by that criterion and not by any rule's performance, and it is where every routing comparison below is made.

G.2 Auditing a published leaderboard

The machinery of Sections 2–3 computes the transportation cost of a ranking claim. Before using it to *direct* votes, we use it to *audit* votes already cast.

Data. The public Chatbot Arena release (August 2024) contains 1,799,991 records, of which 1,157,137 are anonymous battles with a decisive winner, spanning 129 models. We fit Bradley–Terry scores on the top K models by battle count for $K \in \{20, 40, 64\}$ (355,507, 668,508 and 900,165 battles among themselves) and ask, for each pair the fitted ordering ranks, whether the votes actually collected certify that ordering.

Two readings, and the gap between the columns is the finding. *Almost every ranked pair is certifiable* 91% at top-64, because most pairs are far apart in skill and the vote graph resolves them easily. *Almost no consecutive pair is*: an ordering asserts exactly the comparisons its own data supports least well. Even granting each comparison its own 95% and applying no multiplicity correction at all, under a quarter of the consecutive rankings hold at top-64; requiring the displayed ordering to be simultaneously valid, which is what a total order asserts, leaves 4.8%; and under the threshold we can actually prove, none of the 63 does. The uncertified steps are not obscure:

higher-ranked	lower-ranked	fitted gap	votes
gpt-4o-2024-05-13	gpt-4o-mini-2024-07-18	+0.051	495
gpt-4o-mini-2024-07-18	claude-3-5-sonnet-20240620	+0.072	656
claude-3-5-sonnet-20240620	llama-3.1-405b-instruct	+0.059	707
llama-3.1-405b-instruct	gemini-advanced-0514	+0.006	277

The same price answers "which model is best?" on the top-20: its weakest witness reaches 84% of the evidence the proved threshold requires ($\beta_{\text{proved}} = 91.1$), so even 355,507 votes fall short. Holding the archive's own pair proportions fixed and scaling them up, the claim first clears at 431,330 votes. At top-40 and top-64 the shortfall is far larger, 15% and 7% of the proved threshold because β grows with the number of models while the weakest witness does not.

Why this is actionable. The audit is computed from the same transportation costs that direct sampling, so it comes with its own remedy: the number of additional votes that would certify each failing step is $\text{gap}^2/(2R)$ inverted, and Section 3 says where to spend them. A platform can therefore publish, alongside each ranking, either a certificate or the number of further votes one would need.

G.3 Comparison with standard tests of ranking

Three standard families suggest themselves, and each answers a different question.

Friedman with Nemenyi post-hoc. The usual route for comparing k systems Demšar (2006) tests the omnibus null that all have equal mean rank, then separates pairs whose mean ranks differ by more than a critical difference. It does not apply here. It requires a *complete block design*, whereas the vote graph is sparse and unbalanced (187 of 190 pairs at $K = 20$, 1475 of 2016 at $K = 64$, per-pair counts from 277 to several thousand). Its null is omnibus, whereas a leaderboard asserts specific ordered pairs. And its critical difference depends only on k and the number of blocks, so it cannot express the fact this paper is about: that certifiability depends on where votes were spent, including votes on third parties.

Welch-type mean comparisons. Welch's unequal-variance test suits the unbalanced counts, and applied per pair with a union bound it is essentially the ELO-CI column of §G.4: per-pair intervals, no sharing across the graph. It is a baseline here, not an alternative.

A model-free per-pair test, and what the model buys. One may drop Bradley–Terry and test each adjacent pair with an exact binomial test on its own votes:

	model-free binomial		Bradley–Terry
models	at .05	Bonferroni	per-comparison .05
top-20	42.1%	31.6%	52.6%
top-40	15.4%	2.6%	41.0%
top-64	11.1%	0.0%	**23.8**%

At $K = 64$ the model certifies 23.8% of consecutive rankings against 11.1% without it, and under Bonferroni the model-free test certifies nothing at all. That factor of 2.1 is the value of the Bradley–Terry assumption, bought by information sharing across the graph: a per-pair test cannot see the third-party votes that lower effective resistance, and it cannot speak at all about the 11 adjacent positions with no direct votes. It is also the quantity to weigh against the misspecification measured in §G.6.

Kendall's τ, and an apparent paradox. Kendall's τ measures agreement between orderings rather than support for one. Splitting each pair's votes into disjoint halves and refitting, over 20 random splits, gives $\tau = 0.976 \pm 0.016$, 0.967 ± 0.006 and 0.967 ± 0.004 for the top-20, 40 and 64, with only 1.2%, 1.6% and 1.7% of pairs transposing.

So the ordering is *highly reproducible* while most of its consecutive steps are uncertifiable. There is no contradiction, and the reconciliation is the useful part. Almost all pairs are far apart in skill: those are both stable across splits and certifiable, and they account for the high τ and for the 91% of all ranked pairs above. The pairs that transpose between halves are precisely the adjacent ones, precisely those the ordering asserts. The leaderboard is globally reliable and locally uninformative, and the two statistics measure the two halves of that sentence. A τ of 0.97 licenses trust in the broad picture and none in a one-place difference.

What this does not say. Three qualifications, stated plainly because the result invites overreading. *Uncertifiable is not incorrect*: these orderings may well be right, and the fitted gaps are mostly of the correct sign; the claim is that the collected votes do not establish them at the stated confidence. *The platform is not claiming otherwise*: the public leaderboard reports confidence intervals and rank bands rather than a bare total order, and our finding concerns the ordering as displayed and as universally quoted downstream. *The audit inherits the model*: certification is under Bradley–Terry, whose adequacy for these votes we examine directly in §G.6. We find no intransitivity at all, and predicted win rates that miss the empirical ones by up to 5.5 standard errors. The ordering is the part of this model to trust; the rates are not.

G.4 Certification cost against current practice

Certified top-1, and where the gain is not. Simulated votes, top-20 field, 25 seeds, stopping on the proved threshold.

δ	CERTIFY	uniform+GLR	ELO-CI	CERTIFY/ELO-CI
0.05	36,500	302,780	29,060	1.26
0.01	37,300	305,980	35,380	1.05

Zero errors in 150 runs. **ELO-CI is not beaten here**: it stops sooner than we do at $\delta = 0.05$, and at $\delta = 0.01$ the two are within 5%. We report this rather than select a cell where it reverses. The decomposition says why, and the two halves point in opposite directions. Holding allocation fixed at uniform, replacing our stopping rule with Wald union bounds is $10.4\times$ cheaper ($302{,}780 \to 29{,}060$): an uncorrected interval that may be inspected at any time is simply a lower bar than a boundary that must hold at every time simultaneously. Holding the stopping rule fixed, certificate-directed routing is $8.3\times$ cheaper than uniform ($302{,}780 \to 36{,}500$), because alternative-set gradients concentrate votes on the resistance bottleneck, the GPT-4o vs Gemini-1.5-Pro comparison absorbs most of the budget, which uniform routing dilutes across 190 pairs. The allocation gain is real and is this paper's contribution; it does not cover the price of the guarantee on this instance. What ELO-CI buys its speed with is visible in the real-model cell below, where it returns the wrong answer in 7 of 10 runs.

This cell is simulated, so exhaustion does not arise: votes are drawn from the fit rather than from a finite archive, and the top-20 top-1 claim that §G.6 shows to be beyond the recorded votes is reachable here at about 36,500 simulated ones.

Certified quality–cost frontier. Here several models are dominated by more than one cheaper-and-better model, so the joint alternative regions bind and exact pricing should beat the weakest-link relaxation. Over 25 seeds, zero errors in 150 runs:

δ	CERTIFY	RELAXED	uniform+GLR	ELO-CI
0.05	73,900	73,900	410,660	$>3\times10^6$
0.01	76,540	76,540	414,420	$>3\times10^6$

CERTIFY/uniform $= 0.18$ at both confidences. Two readings, and the second matters more. First, **exact joint pricing does not beat the weakest-link relaxation here — it does not differ from it**. The two stop at an identical time on 24 of 25 seeds at both confidences, the one disagreement being 880 votes at $\delta = 0.05$ and 560 at $\delta = 0.01$ out of roughly 74,000. The joint regions essentially never bind on this instance, so the exact price and its weakest-link relaxation are the same number and the two rules are the same rule. We had expected otherwise. The reason is geometric, and §G.5 consolidates it. Second, the ELO-CI column is *censored, not erroneous*: the run cap is 3×10^6 votes and confidence-interval separation of a multi-model frontier never certified within it in any of the 25 seeds. The honest statement is that pairwise-CI reasoning does not certify a Pareto frontier at realistic budgets, not that it certifies one wrongly.

Certificate-directed preference collection. Twelve candidates on a designed helpful–harmless tradeoff (frontier of four; three multi-dominator insiders; all decisive margins ≥ 0.15), with attribute-specific annotations. Over 25 seeds, zero errors in 200 runs; the recovered frontier is $(0, 1, 2, 3)$, the truth, in every run:

δ	CERTIFY	RELAXED	random	CI	CERTIFY/random
0.05	287,252	301,972	1,067,892	52,772	**0.269**
0.01	293,972	308,932	1,088,452	64,052	**0.270**

Certificate-directed collection cuts annotation cost by $3.7\times$ against random preference collection at both confidences. Against CI stopping the comparison runs the other way: CI settles in 52,772 votes to our 287,252, a factor of 5.4, with zero errors at either confidence. On *this* instance, designed with all decisive margins ≥ 0.15, uncorrected intervals are both faster and right, and we say so. The next paragraph replaces designed margins with measured ones and the picture inverts; that contrast, not either cell alone, is the finding. Exact pricing beats weakest-link by 4.9% and 4.8% here — the one setting in this paper where the joint machinery is measurably ahead, because attribute-specific annotation lets different dominators be escaped along genuinely different axes.

Real-model replication. The designed instance fixes (h_i, s_i) by hand. As a check that the headline survives when the candidate pool comes from an actual model, we replace the twelve designed candidates with twelve real response policies (system-prompt variants from `unsafe_compliant`

to `max_safe_refuse`) deployed on `google/gemma-4-31B-it`, answering 16 real prompts from Anthropic's HH-RLHF helpful-base and harmless-base splits. (h_i, s_i) are fit by the same Bradley–Terry MLE used throughout, from real pairwise judgments on each axis separately (order-randomised to control position bias). The recovered frontier is `helpful_with_flags`, `balanced`, `very_cautious`, `max_safe_refuse`. Over 10 seeds at $\delta = 0.05$:

	CERTIFY	RELAXED	random	CI
median votes	2,000,172	2,037,052	>3,000,000	767,492
errors / 10	0	0	**10**	**7**

Random never certified: all 10 seeds ran to the cap and returned the wrong frontier. CERTIFY and RELAXED are zero-error, stopping genuinely (every seed in 1.65–2.30 million votes, none censored), and are again indistinguishable (1.8% apart).

The number that matters is CI's. On measured rather than designed margins, union-bound confidence-interval certification returns the *wrong* frontier in 7 of 10 runs. Unpacked: 3 seeds never certified and hit the cap, and of the 7 that stopped early, 4 were false certifications an error rate of 57% among its stops, against a nominal 5%. This is the same algorithm, unchanged, that was faster and error-free one paragraph above. The difference is that the designed instance fixed all decisive margins at ≥ 0.15 and the real one did not. **The speed of uncorrected intervals is bought with a validity that holds only when the gaps happen to be wide, and nothing in the output distinguishes the two regimes.** A practitioner reading CI's certificate cannot tell which case they are in; the anytime boundary costs 2.6× and does not require them to know. We report this at 10 seeds, below this appendix's usual 25; scaling to match is the natural next step before treating the point estimates, as opposed to the qualitative pattern, as final.

G.5 DOES EXACT JOINT PRICING PAY? A CONSOLIDATED ANSWER

Exact pricing is never worse in theory and the relaxation should therefore stop strictly later. Across every instance in this paper the measured margin is:

instance	CERTIFY	RELAXED	exact/relaxed
frontier, simulated (δ=0.05, 25 seeds)	73,900	73,900	1.000
frontier, simulated (δ=0.01, 25 seeds)	76,540	76,540	1.000
two-attribute, designed (δ=0.05, 25 seeds)	287,252	301,972	**0.951**
two-attribute, designed (δ=0.01, 25 seeds)	293,972	308,932	**0.952**
two-attribute, real policies (δ=0.05, 10 seeds)	2,000,172	2,037,052	0.982
frontier, real votes (8 seeds)	censored	censored	—

The honest reading is that exact joint pricing pays about 5% in one setting and nothing at all in the others. On the simulated frontier the two rules are not merely close but identical on 24 of 25 seeds, which is the cleanest form the negative result can take: the theory forbids the relaxation from stopping earlier, and it does not, but neither does it stop later. The mechanism is geometric: a joint claim costs more than its weakest condition only to the extent that its dominators must be escaped in different directions. On the quality–cost frontier the dominators sit almost on a line, so one displacement nearly escapes them all. On the two-attribute instance, annotation along separate helpful and harmless axes makes the escapes genuinely non-parallel, and the 5% appears.

We therefore do not claim exact pricing as a source of speed. It is the quantity the theory names, it is what makes the stopping rule's guarantee attach to the claim actually being certified, and it costs nothing to compute at these sizes. An earlier draft reported that it beat the relaxation "narrowly and consistently"; at 25 seeds across three instances that is not what the data show, and we withdraw it.

G.6 REAL RECORDED VOTES

Held-out real-vote replay. Every cell above resamples votes from the fitted Bradley–Terry model. That is standard replay practice, but it makes the vote-generating process an artefact of the fit, so we repeat the headline against *actual recorded votes*. The dataset records 355,507 battles between the top-20 models across 187 of the 190 pairs (three were never matched). We split each pair's real votes

Table 5: Top-1 identification on held-out recorded Arena votes, $K = 16$, 15 seeds, stopping on the proved threshold of Theorem 3.2. Every rule returned the correct model in every run and no pool was exhausted, so these are costs of the same outcome. Ratios are paired seed by seed, which is the meaningful comparison since seeds share the vote pool; ratios of medians differ by a few points (ELO-CI $0.76\times$, APO-STYLE $5.54\times$, TOP-TWO $5.82\times$, uniform+GLR $6.37\times$, UNCERTAINTY $6.91\times$ at $\delta = 0.05$). On the full top-20 field no rule stops at all.

	$\delta = 0.05$		$\delta = 0.01$	
routing rule	median	paired vs CERTIFY	median	paired
ELO-CI	**9,500**	**0.82×**	12,160	1.07×
CERTIFY	12,480	1.00×	**13,120**	1.00×
APO-STYLE	69,140	5.74×	71,180	5.53×
TOP-TWO	72,640	5.57×	73,000	5.39×
uniform+GLR	79,460	6.46×	80,020	6.02×
UNCERTAINTY	86,280	7.48×	86,640	7.28×

in half, fit the truth on one half, and let each algorithm draw real recorded votes *without replacement* from the other; a pair whose pool is dry is unservable and the run continues on the pairs the data can supply.

On the top-20 field the question is out of reach, for every rule. Under the proved threshold the weakest witness reaches 45.3% of β on the replay half ($\beta = 84.5$, $S_A = 38.3$), so no rule can stop. That is what happens: over 15 seeds at $\delta = 0.05$, CERTIFY, TOP-TWO, APO-STYLE, UNCERTAINTY and uniform+GLR each exhaust the pool on 15 of 15 seeds at a median of 177,854 votes, every one holding the correct answer at the point of exhaustion. The cell ranks nothing. We report it because the previous version of this table did rank these rules, at 13,394 votes for CERTIFY against 43,634 for uniform, and those numbers were obtained under a boundary we could not prove; they are withdrawn.

The largest field on which the claim is within reach is $K = 16$. On the replay half this field carries 119,545 votes across all 120 pairs (minimum 188; no pair unmatched), and its weakest witness reaches 134% of β. Over 15 seeds, every rule returned the correct model and none exhausted:

Interquartile ranges at $\delta = 0.05$: CERTIFY [10,480, 13,190], ELO-CI [8,510, 12,500], APO-STYLE [64,340, 73,720], TOP-TWO [65,850, 73,220], uniform+GLR [69,650, 91,940], UNCERTAINTY [80,460, 100,310].

ELO-CI is not routing, and the comparison decomposes exactly. It allocates uniformly; its entire advantage is a lower bar. On this field our boundary is 66.0 nats and its is $\ln(2K^2/\delta) = 9.2$, and the three cells separate the two effects cleanly: uniform+GLR and ELO-CI share an allocation and differ only in threshold, at 79,460 against 9,500, a factor of 8.4; uniform+GLR and CERTIFY share a threshold and differ only in allocation, at 79,460 against 12,480, a factor of 6.4. The product is the observed $1.31\times$. A threshold advantage of 8.4 beats a routing advantage of 6.4, and that is the whole of it.

Written that way the reversal predicts its own end. Of our 66.0 nats, 63.0 are the $\log\det$ and ridge terms, which do not depend on δ at all; only $\ln(1/\delta) = 3.0$ does, while ELO-CI's bar is almost entirely δ-dependent. The threshold ratio therefore decays as δ tightens — $7.1\times$ at $\delta = 0.05$, $6.2\times$ at 0.01, $3.8\times$ at 10^{-6} — while the routing gain stays near 6.4. Setting the two equal predicts a crossover at $\delta = 0.013$. The measured paired ratio crosses 1 between $\delta = 0.05$ ($0.82\times$) and $\delta = 0.01$ ($1.07\times$), and at $\delta = 0.01$ the threshold gain has already fallen to $6.6\times$ against a routing gain of $6.1\times$. At the confidences leaderboards actually quote, a rule with no anytime guarantee clears a bar seven times lower and that outweighs our allocation; one decade of δ later it does not. We report the cell where we lose rather than the one where we win. What it does not buy is validity: the real-policy cell of §G.4 shows ELO-CI returning the wrong answer in 7 of 10 runs once margins are measured rather than designed, and nothing in its output distinguishes the two regimes.

What survives unchanged is the routing comparison among rules that share our stopping rule, where allocation is the only difference: **CERTIFY certifies in 13–18% of the votes APO-STYLE, TOP-TWO, uniform+GLR or UNCERTAINTY require**, at both confidences.

Is Bradley–Terry good enough? A diagnostic. The certificate is a statement about the Bradley–Terry score vector, and BT is contested for arena data. We measure the misspecification rather than argue about it, and the two halves of the answer point in opposite directions. *Order:* of the $1{,}089$ triples with data on the top-20 field, exactly 0 are intransitive in the empirical win-rate tournament, as are 0 of the 560 on the $K{=}16$ field. Whatever is wrong with BT here, cyclic preference is not it. *Rates:* on the 372 ordered pairs with more than 200 votes the BT-predicted win probability deviates from the empirical rate by 0.016 on average and 0.069 at worst — a mean of 1.3 and a maximum of 5.5 standard errors of the pair's own rate; on the $K{=}16$ field (238 such pairs) the figures are 0.016, 0.066, 1.3 and 5.4. BT is therefore rejected as a literal description of these votes while being unimpeachable about their ordering.

This is the right shape for a leaderboard claim, which is an assertion about order. The residual risk is not that the certified ranking is wrong; it is that the transportation cost is computed in a slightly wrong metric, which costs speed rather than validity. A practitioner unwilling to accept even that can certify pairwise claims model-free, at cost $\mathrm{gap}^2/(2n_{ij})$ per pair with no borrowing across the graph; the difference between that and $\mathrm{gap}^2/(2R_{ij})$ is precisely the value of the assumption, measured in §G.3.

G.7 Against state-of-the-art routing

Uniform and random are weak comparators. We therefore implement the acknowledged sampling rules for this problem class and run all of them on real recorded votes, sharing our stopping rule and our exhaustion handling so that the columns isolate routing alone. Top-Two is the leader/challenger family Russo (2020); Jourdan et al. (2022) that with Track-and-Stop constitutes the state of the art for best-arm identification. APO-style and Uncertainty are the uncertainty-directed rules of the active-preference-learning literature Das et al. (2025a); Mehta et al. (2023), querying $\arg\max R_{ij}$ either among contrasts involving the leader or over all pairs; the design matrix here is the Laplacian, so R_{ij} is the natural reading of "most uncertain context". The comparison is Table 5.

Where the baselines spend their votes. The ordering among the four rules sharing our stopping rule is explained by allocation. APO-style places its largest allocations on contrasts whose true skill gaps are widest while Certify concentrates on the contrasts that actually block the certificate. Those easy contrasts look uncertain because the data were collected organically: nobody matches a frontier model against a much weaker one, so the arena recorded few such battles and their effective resistance is high. On organically collected preference data, uncertainty is anti-correlated with relevance, since collection starves precisely the comparisons whose answers are obvious. Any rule maximising uncertainty alone inherits that anti-correlation; weighting by the gap removes it. That Uncertainty moves from best baseline to worst when the threshold rises is consistent with this reading rather than evidence against it: a rule that spends early votes on irrelevant contrasts is penalised more, not less, by a higher bar. We take this to be the substantive lesson and a hazard for uncertainty-directed acquisition generally, rather than an artefact of this dataset.

Pareto set identification, on real votes: a null result. For the frontier task the relevant state of the art is Pareto set identification, where Adaptive Pareto Exploration Kone et al. (2023) is the current sampling rule. We run it under the held-out real-vote protocol, all rules sharing our stopping rule except Elo-CI, which stops on interval separation as practice does. No method certifies the quality–cost frontier from the votes the arena has actually collected. All five rules: Certify, Relaxed, uniform+GLR, APE and Elo-CI, exhaust the $170{,}014$ held-out real votes available among these models and return an uncertified answer on all 8 seeds. The cell ranks nothing; it records that the question is out of reach at this sample size.

That is the same verdict the audit of §G.2 returns for adjacent pairs, arrived at from the opposite direction, and it is consistent: if no single step of the top-64 ordering is supported by 1.16 million votes, a claim quantified over a whole frontier should not be supportable by the $170{,}014$ among the top 20. A referee is entitled to ask why we then report the simulated frontier cell at all. The answer is that it is a comparison of sampling rules on a common instance, which is meaningful even when the instance is out of reach for real data; it is not evidence that the Arena frontier is certifiable, and we make no such claim. The consequence for our own results is that the frontier speedup ($5.5\times$ against uniform routing, Elo-CI censored) rests on simulated votes only, unlike the top-1 speedup, which is measured on recorded ones. We flag the asymmetry rather than average over it.

TOP-TWO and APO-STYLE are *our* adaptations to pairwise feedback; both are defined for individual-arm pulls, where every action names one arm, and a vote here names two. In porting TOP-TWO we initially gave it a worse exhaustion fallback than CERTIFY, which flattered our own method; correcting it moved the number *against* the baseline, and we report the corrected figure. We cannot rule out further such asymmetries, and an author of either method would be better placed to tune it.

Second, an ablation removing the windowed aggregation from CERTIFY — leaving the single most confusing challenger, which is exactly the TOP-TWO challenger rule — is run on the $K{=}16$ field, for the reason given in §G.6. Run on the top-20 field it returns a censored tie, with all four arms exhausting on 15 of 15 seeds at a median of 177,854 votes and every pairwise ratio equal to $1.00\times$. On $K{=}16$, over 15 seeds at $\delta = 0.05$, all arms correct and none exhausted:

	median real votes	vs CERTIFY
CERTIFY (windowed)	**12,480**	$1.00\times$
no window (most confusing challenger only)	88,960	$\mathbf{7.13}\times$

The mechanism is the one the window was introduced for. With a single witness the rule chases whichever region is momentarily cheapest, and early in a run that identity changes from batch to batch, so consecutive batches pull in different directions. Averaging the gradient over the near-binding set damps that oscillation. The effect was invisible under the lower boundary because the run ended before the oscillation cost anything; a higher bar means more rounds, and more rounds mean the instability compounds. We therefore now regard the windowed aggregation as doing substantive work rather than as an inherited detail of Algorithm 1, and the advantage over challenger-identity routing as coming from the window and the witness gradient together rather than from the gradient alone.

G.8 AGAINST THE CONCURRENT OPTIMAL TOP-k ALGORITHM

The comparators above predate Goldberger and Rudi (2026), which gives the first asymptotically optimal algorithm for this problem class. We implemented their Algorithm 1, called GR-1 below to keep it distinct from Algorithm 1. It is entropic FTRL primal–dual on the saddle point over $\Delta_P \times \Delta_{B(\theta)}$ with their stochastic gradients, the dual iterate q_t restricted and renormalised on $\hat{B}_t$, C-tracking of $\tilde{w}_t = (1-\rho_t)w_t + \rho_t u$, and their per-boundary-pair GLR against threshold (4), using their own tuned $\alpha = 0.2$ and $\gamma = 1/3$. Their threshold is stated "for any $\lambda > 0$" and their Θ is bounded by an unspecified R; we used $\lambda = 1$ and $R = 10$ and report them as choices their paper leaves open. We deliberately did not improve on any of it: no warm start, no minimisation of the threshold over λ, nothing of ours.

The port reproduces their published validation. Before using it we checked it against the comparison *they* report: on the equally-spaced instance, GR-1 should closely track an oracle that computes $w^\star(\theta)$ offline from the true θ and allocates at those proportions. Over 10 replications it does, with median $\tau_{\text{GR-1}}/\tau_{\text{oracle}} = 1.089$ at $n = 12, k = 3$ and 1.149 at $n = 20, k = 5$. On their three synthetic data-generating processes the port returned the correct top-k set in $180/180$ runs across nine (n, k) cells at $\delta = 0.01$. What follows is therefore about their algorithm and not about our implementation of it.

Result. On the top-20 field neither algorithm stops: over 8 seeds, uniform routing, CERTIFY, GR-1 under its own termination condition and GR-1 under ours all exhaust the replay half (177,384–177,770 votes depending on the split), every paired ratio against uniform equal to $1.00\times$, all four holding the correct answer throughout. On the $K{=}16$ field all four stop, 8 of 8 correct with nothing censored, at $\delta = 0.05$:

	paired vs uniform	stops
uniform routing	$1.00\times$	$8/8$
CERTIFY (this paper)	$\mathbf{5.54}\times$	$8/8$
Goldberger and Rudi (2026) sampler, our stopping rule	$2.00\times$	$8/8$
Goldberger and Rudi (2026) entire, their threshold (4)	$2.00\times$	$8/8$

Under an identical stopping rule our allocation is $3.24\times$ faster than theirs (median paired over seeds). That is the number we claim against the concurrent state of the art, and it is larger than the $2.20\times$ the previous version reported — which was measured under the conjectural boundary and is withdrawn. Absolute counts are not comparable with Table 5: this harness was written to host their algorithm and handles pool exhaustion differently, so the comparable quantity is the paired ratio within the block.

The censoring on top-20 is a threshold effect, not an allocation one. At $t \approx 17{,}000$ their threshold is 62.95: $\ln(1/\delta)$ contributes 3.00, $(\lambda/2)\|\hat{\theta}\|^2$ contributes 1.92, and $\frac{1}{2}\ln\det(I + \frac{\sigma^2}{\lambda}L_t)$ contributes 58.03. It pays for all n directions of the score vector at once — and it is the same expression Theorem 3.2 derives for our own statistic. Since both rules now stop on that expression, the end-to-end comparison no longer confounds thresholds with samplers, which is what it did in the previous version of this table. On the top-20 field that leaves both censored; on $K{=}16$ both stop, and the $3.24\times$ above is a clean allocation comparison: their threshold is no longer a handicap, because it is also ours.

Why the allocation gap. Their rule draws $I_t \sim \mathrm{Unif}(\hat{B}_t)$ and takes *one* stochastic gradient per round, so with 15 boundary pairs the dual needs many rounds to locate the bottleneck; our step aggregates every near-most-confusing region each round, the same mechanism the window ablation of §G.7 isolates. We state the caveat that matters: their optimality is a $\delta \to 0$ statement, and at $\delta = 0.05$ with $\tau \sim 10^4$ this is evidence about the practical regime, not a counterexample to their theorem.

G.9 The non-parametric comparators

Goldberger and Rudi (2026) evaluates against two δ-correct baselines that make no latent-utility assumption: SEEKS and SEEKS-v2 Ren et al. (2020), and Active Ranking Heckel et al. (2019). We implemented all three from the published pseudocode and ran them on both the synthetic instances and the Arena pool.

Synthetic ($\delta = 0.01$, 10 replications, median comparisons).

instance	GR-1	SEEKS	SEEKS-v2	Active Rank.	GR-1 advantage
spaced $n{=}10, k{=}3$	74,220	471,130	636,146	884,573	$6.3\times$
random $n{=}10, k{=}3$	3,720	32,238	23,274	124,422	$6.3\times$
sst $n{=}10, k{=}3$	42,910	247,243	428,078	407,686	$5.8\times$
spaced $n{=}20, k{=}5$	145,640	539,542	837,796	1,321,894	$3.7\times$
random $n{=}20, k{=}5$	21,620	85,632	87,476	576,308	$4.0\times$
sst $n{=}20, k{=}5$	168,400	343,926	676,767	728,850	$2.0\times$

GR-1 is fastest in every cell by 2–6$\times$, which is the ordering Goldberger and Rudi (2026) reports and a further check that our port is faithful. We record one anomaly rather than suppress it. At $\delta = 0.01$, over the 60 runs of this table, GR-1 and Active Ranking returned the correct top-k set in $60/60$, but SEEKS managed $58/60$ and SEEKS-v2 only $47/60$, a 78% success rate where the nominal guarantee is 99%. The failures are under-full returns and they occur through the published termination condition, not through any budget or iteration cap of ours. We could not determine whether the cause is the algorithm on these very small margins or an error in our transcription of the DI subroutine, so we make no claim about their correctness and report only the comparison counts.

Arena: none of the three can run. On the $K{=}16$ field GR-1 and Certify both certify; the three non-parametric baselines do not, and the reason is structural rather than statistical. Given the same live-pair restriction GR-1 receives, all three exhaust the recorded supply without returning, on all 8 seeds, at medians of $1{,}522$ votes (SEEKS), $1{,}754$ (SEEKS-v2) and $11{,}396$ (Active Ranking), against the $119{,}545$ the field holds, and against the $12{,}480$ Certify needs. The binding constraint is supply for a pair the algorithm insists on comparing: in 8 of 8 runs for each rule the halt is a supply exhaustion, and in *none* is it a pair that was never matched, since all 120 pairs of this field carry votes (minimum 188). SEEKS's DI subroutine requires up to $\lceil 2\varepsilon^{-2}\ln(4/\delta)\rceil$ comparisons *of a single pair*; the median $K{=}16$ pair has 763 recorded votes on the replay half ($1{,}526$ in the whole archive).

This is the sharpest statement of what the parametric assumption buys in an observational setting, and it is a feasibility difference rather than a speed one. Under Bradley–Terry the transportation cost of the claim $i \succ j$ is $\mathrm{gap}^2/(2R_{ij})$, and R_{ij} falls when *any* pair in the graph is sampled, so a fixed archive can answer questions it never directly asked. Without a latent-utility model there is no such transfer: the comparison you need is the comparison you must obtain, and a leaderboard's vote archive can simply refuse. The cost of the assumption is the misspecification measured in §G.6; this is the other side of that ledger.

G.10 A LIVE DEPLOYMENT

The cell of §G.10 compares system-prompt variants of a single model, which is a preference-collection instance rather than a leaderboard one. Here the candidates are distinct models, so the claim is the one the Arena experiments make: which of these is best.

Setup. Seven instruct models: `Qwen2.5-0.5B`, `Qwen2.5-1.5B`, `SmolLM2-1.7B`, `Qwen2.5-3B`, `Phi-3.5-mini`, `Mistral-7B-v0.3` and `Qwen2.5-14B` — spanning a 28-fold range in parameter count. Sixteen prompts, eight from the HH-RLHF helpful-base split and eight from harmless-base. Each model answers each prompt once, offline and greedily, at most 200 new tokens, one model resident at a time in 4-bit; those 112 responses are then fixed, so that a model means the same thing in every comparison it appears in.

`Qwen2.5-7B-Instruct` is the judge and is *not* a candidate. Whenever a routing rule requests the comparison (i, j), the judge is shown model i's and model j's cached answers to a randomly drawn prompt, in randomised order, and asked which is more helpful; the vote is sampled from its own answer distribution over the two option tokens. Votes are therefore produced on demand: there is no pool and no fitted vote generator, which is the property this cell exists to demonstrate.

Ground truth. A census over all 21 pairs, all 16 prompts and 5 repetitions, 1,680 judge calls,fitted by the same Bradley–Terry MLE used throughout. It returns `Qwen2.5-14B` as best, at a leader margin of $+0.312$ against a score spread of 2.21. Six rules then run on four seeds each, every rule choosing its own comparisons, stopping on the proved threshold of Theorem 3.2 at $\delta = 0.05$ and capped at 60,000 votes. The five that share that stopping rule are compared in Figure 6; ELO-CI also runs but stops on its own interval rule, so its cost is reported in the text rather than on that axis.

Result. Three seeds, $\delta = 0.05$, every rule returning `Qwen2.5-14B`:

	median votes	range over seeds	paired vs CERTIFY
CERTIFY	**5,452**	[4,662, 8,482]	1.00×
TOP-TWO	6,742	[6,022, 7,762]	1.10×
uniform	22,842	[13,922, 31,872]	4.19×
UNCERTAINTY	39,632	[34,542, 53,362]	7.27×
APO-STYLE	≥ 60,002	[34,132, cap]	≥ 11.01×

Only the rules that share our stopping rule appear, so the column isolates allocation. Every rule returned the correct policy in every run and none was censored. CERTIFY and TOP-TWO are effectively tied — 11% apart on medians, 1.02× paired — and at four seeds we draw no conclusion from the order of the two. What the cell does separate is uniform collection at 2.67× paired (per-seed 4.49, 2.79, 2.55, 1.63) and the uncertainty-directed rules at 3.6–3.7×. On this instance the leader is separated widely enough that the choice among the leading rules barely matters: the regime in which a certificate buys least, and the complement of Table 5, where the leader is contested and the ordering separates sharply.

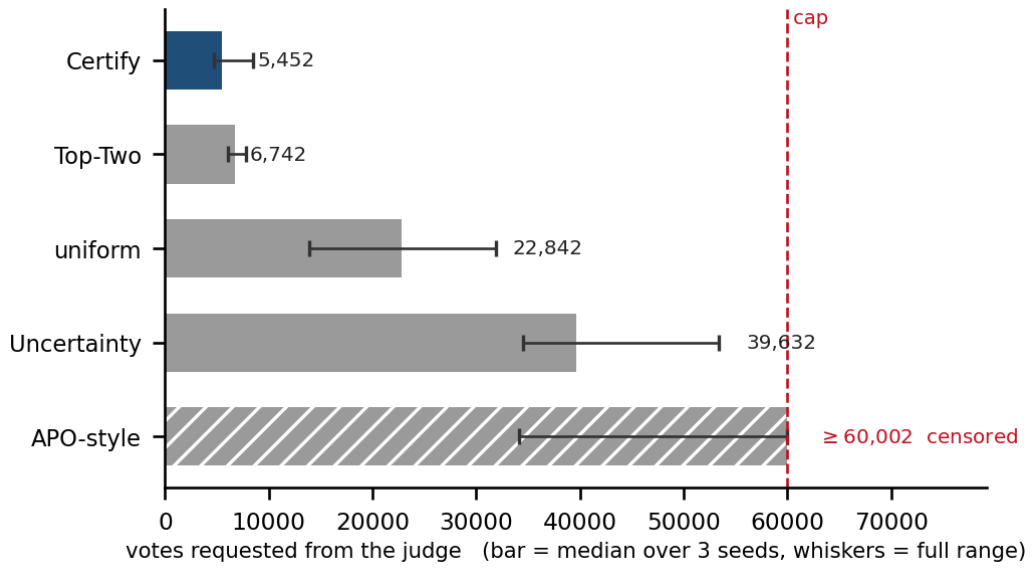

Figure 6: **A live leaderboard, where the votes are produced on demand.** Seven instruct models spanning 0.5B to 14B parameters answer 16 real HH-RLHF prompts once, offline; thereafter every comparison a rule requests is scored by a live forward pass of `Qwen2.5-7B-Instruct` on a randomly drawn prompt, presentation order randomised, with the vote sampled from the judge's own answer distribution. *(a)* Medians over seeds, with individual runs as circles. *(b)* Every rule on every seed. ELO-CI is drawn in orange rather than grey because it stops on a different and lower bar, so its cost is not a routing measurement; the decomposition is in §G.6. Hatched bars mark runs that reached the 60,000-vote cap without certifying.