BALANCING STRUCTURE AND STABILITY IN LEADERBOARDS WITH QUANTAL RESPONSE EQUILIBRIUM

Anonymous authors
Paper under double-blind review

ABSTRACT

Public leaderboards shape how AI models are perceived and adopted. Yet defining a ranking from preferences is nontrivial, as different aggregation principles can favor different models. In practice, not every pair of models is compared, raising the additional challenge of predicting missing entries while retaining competitive structure. Elos scalar model restricts comparisons to additive logits, excluding potentially non-transitive structure. On frontier leaderboards, we show that adding a synthetic model that loses to everyone dethrones Elos leader, despite that leader still beating every rival and all original comparisons remaining unchanged. Maximal lotteries (MaxL), an alternative from probabilistic social choice, capture arbitrary tournament structure, but require a complete comparison matrix. Their outputs can change substantially across resamples. To reconcile structure and stability, we decouple leaderboard construction into comparison-matrix estimation and ranking. For estimation, we derive error bounds for low-rank completion that account for structural approximation, regularization, and finite-sample error. These bounds quantify the statistical cost of richer representations. For ranking, we analyze quantal response equilibrium (QRE), an entropy-regularized counterpart to MaxL that yields smooth, payoff-based scores. We derive temperature-dependent stability bounds and approximate guarantees for *Condorcet consistency* and *independence of irrelevant alternatives*. Empirically, QRE preserves the incumbent order in the insertion example and retains its leader in nearly all resamples.

1 INTRODUCTION

Evaluating increasingly capable models on open-ended tasks often relies on pairwise comparisons of their outputs. Preference-based leaderboards such as LMSYS Chatbot Arena (Chiang et al., 2024) and Datapoint Image Bench (Datapoint AI, 2026) aggregate these comparisons into model rankings that guide model development and adoption. Constructing such a leaderboard involves two choices: how to estimate a complete comparison matrix from finite, incomplete observations, and how to aggregate that matrix into a ranking. The estimation model determines which competitive relationships between models can be represented. Different ranking rules can favor different models on the same matrix and respond differently to estimation error. This paper asks: *How can a leaderboard respect competitive structure while remaining stable under sampling noise?* We analyze estimation and ranking separately to characterize how accurately the comparison matrix can be estimated and how the choice of ranking rule affects the reported outcomes and their stability.

Predicting unobserved comparisons requires structural assumptions. Elo-style leaderboards based on the Bradley–Terry (BT) model (Bradley & Terry, 1952) use a single scalar rating per model to predict pairwise win rates and rank models. BT implies additive assumption on preferences, excludes potentially non-transitive structure regardless of how many votes are collected. To represent non-transitive structures with adjustable capacity, we adopt the existing low-rank DISC-k representation (Bertrand et al., 2023). Increasing its capacity expands the representable structure, but can also increase the sensitivity to sampling noise. We analyze this trade-off by separating structural approximation from regularization and sampling error, and deriving finite-sample bounds on matrix estimation error.

Once the comparison matrix has been estimated, constructing a ranking still requires deciding how to aggregate pairwise preferences. The classic Borda–Condorcet debate concerns this choice: Borda

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

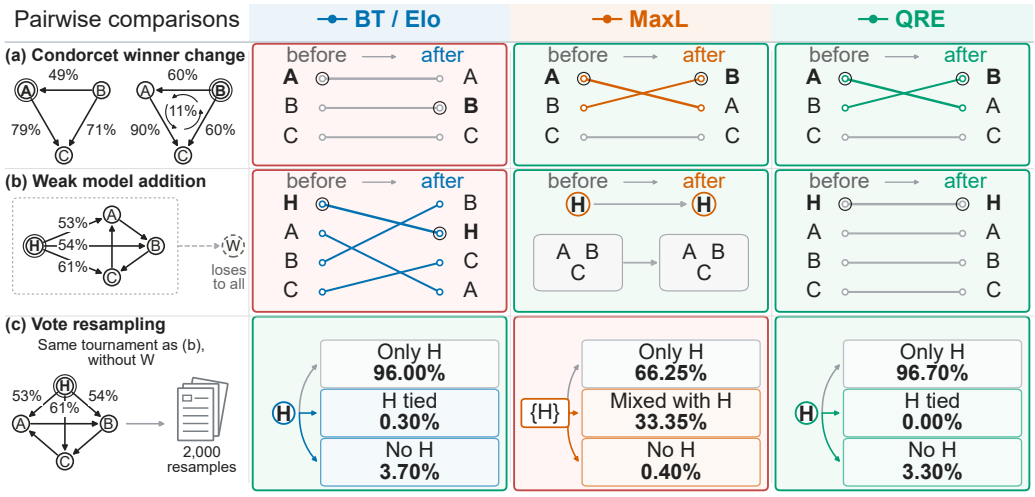


Figure 1: **Ranking rules respond differently to structural changes and sampling noise.** (a) A cyclic perturbation changes the Condorcet winner from A to B while preserving average win rates, leaving BT's ranking unchanged. (b) Adding a model that loses to every incumbent changes BT's leader despite unchanged incumbent comparisons. (c) Percentages summarize 2,000 independent bootstrap resamples of the tournament in (b) before insertion, with 200 votes per pair. Red highlights undesired outcomes; double rings mark Condorcet winners.

rewards average win rate against equally weighted opponents, whereas Condorcet consistency requires a model that beats every rival to rank first (marquis de Condorcet, 1785). Figure 1(a) illustrates their disagreement. Shifting win rates around a triangle preserves every model's average win rate but makes B beat every rival. BT follows the Borda order under mild conditions (Siththaranjan et al., 2024), still ranks A first: its advantage over C outweighs its loss to B.

BT ranks by average win rate, so adding an opponent can reorder the existing models. In Figure 1(b), adding a model W that loses to everyone makes BT rank B first instead of H, although all original comparisons remain unchanged and H still beats every rival. This effect also occurs beyond the toy example: on Image Bench, a synthetic all-loser changes the rankings of 25/30 original models and dethrones the original leader without changing any original comparison (Figure 2).

Maximal lotteries (MaxL), an alternative from probabilistic social choice theory, select a distribution over models that no challenger beats on average, with Condorcet and clone consistency (Fishburn, 1984). MaxL therefore selects the new Condorcet winner B in Figure 1(a) and retains the winner H after insertion in (b). These guarantees do not ensure reproducibility from finite data. Figure 1(c) resamples votes from the tournament in (b), without W. The selected MaxL support frequently changes across resamples. On Chatbot Arena, its support changes in 27.45% of bootstrap resamples.

Resampling changes the estimated comparison matrix, so reproducibility depends on the ranking rule's response to it. To control this sensitivity, we analyze quantal response equilibrium (QRE), an entropy-regularized counterpart to MaxL (McKelvey & Palfrey, 1995). QRE scores each model by its expected performance against a shared distribution of opponents. A temperature parameter controls the strength of entropy regularization and thus how strongly this distribution favors higher-scoring models. At positive temperature, both the opponent weights and the scores vary continuously with the comparison matrix. Increasing temperature tightens their worst-case sensitivity bounds, but can relax social-choice guarantees. We quantify this trade-off and derive score-gap conditions for preserving pairwise orderings under estimation error. In Figure 1, QRE follows the new Condorcet winner in (a), preserves the incumbent order in (b), and retains H in 96.7% of resamples in (c).

Our contributions are: **(i) Estimating structure from incomplete comparisons.** For DISC-k completion, we separate structural approximation, regularization, and sampling error and derive finite-sample bounds. These bounds quantify the statistical cost of retaining structure beyond additive ratings (Section 3). **(ii) Balancing competitive guarantees and reproducibility.** We characterize QRE's temperature-dependent trade-off between sensitivity to sampling noise and social-choice

guarantees. We establish conditions for Condorcet consistency and approximate indenpendence to irrelevant alternatives (Section 4). **(iii) Empirical evidence on preference leaderboards.** Recovery experiments show that low-capacity fits perform better with few votes, while higher capacity improves recovery as data grow. On Arena and Image Bench, DISC models non-transitive structure with win-rate prediction accuracy close to BT. QRE preserves incumbent orders in insertion and deletion experiments and retains its leader in nearly all resamples (Section 5).

1.1 Related Work

Preference leaderboards and robustness. Leaderboards for language and image-generation models depends on BT fits to pairwise judgments (Chiang et al., 2024; Datapoint AI, 2026). Prior work examines these rankings' vulnerability to vote rigging (Min et al., 2025; Huang et al., 2025) and sensitivity to preference deletion and candidate removal (Huang et al., 2026; Oyarhoseini et al., 2026).
Representing and estimating comparisons. Hodge decomposition separates pairwise logits into transitive and cyclic components (Jiang et al., 2011; Balduzzi et al., 2018). DISC accommodates both types of structure through low-rank parameterization and supports incomplete comparisons (Bertrand et al., 2023), although sparse data can still favor simpler Elo models for prediction (Tang et al., 2025). Recent work uses a low-rank matrix of task-specific model scores for ranking and uncertainty quantification, assuming additive logits within each task (Li et al., 2026). Our estimation analysis concerns completion of model-by-model logits and builds on low-rank estimation theory (Davenport et al., 2014; Negahban & Wainwright, 2012).
Probabilistic social choice and ranking. MaxL has axiomatic foundations in probabilistic social choice (Fishburn, 1984; Brandl et al., 2016) and applications to agent evaluation (Lanctot et al., 2023). SCO achieves Condorcet consistency with scalar scores at the global optima (Lanctot et al., 2025). Robust lotteries optimize worst-case performance over uncertain margins, with finite-sample regret guarantees for estimated population-mixture weights (Khalaf et al., 2026). We study how matrix error changes QRE lotteries, scores, and their induced rankings.
Preference learning and alignment. BT-based reward learning connects to Borda aggregation (Siththaranjan et al., 2024). Nash learning from human feedback targets KL-regularized equilibria of preference games (Munos et al., 2024). With a uniform reference over finite alternatives, this formulation recovers QRE (McKelvey & Palfrey, 1995; Sokota et al., 2023).

2 Problem Setting and Background

A leaderboard first estimates an underlying tournament from finite votes, then ranks its candidates. Estimation determines which competitive structures are retained; ranking determines how estimated comparisons are aggregated. We describe how pairwise comparisons are represented and estimated, the desired social-choice properties and the four ranking methods we study.

Setup: Throughout this work, we use two representations of the same underlying tournament: *logits* for estimation and *margins* for ranking. For $M \geq 2$ models indexed by $[M] = \{1, \dots, M\}$, let P_{ij} denote the win rate of model i against j, then $P_{ji} = 1 - P_{ij}$ and $P_{ii} = 1/2$. The logit matrix $\Theta_{ij} = \log[P_{ij}/(1 - P_{ij})]$, for $0 < P_{ij} < 1$, and the margin matrix $A_{ij} = 2P_{ij} - 1$ are antisymmetric and satisfy $P_{ij} = \sigma(\Theta_{ij})$ and $A_{ij} = \tanh(\Theta_{ij}/2)$, where $\sigma(t) = (1 + e^{-t})^{-1}$. Throughout, we use stars $(\Theta^\star, A^\star)$ to denote true population quantities and hats $(\widehat{\Theta}, \widehat{A})$ denote their empirical estimates. For matrices, we use $\|\cdot\|_F$, $\|\cdot\|_{\mathrm{op}}$, and $\|\cdot\|_*$ for the Frobenius, spectral, and nuclear norms, respectively, and write $\|X\|_{\max} = \max_{i,j} |X_{ij}|$. For vectors, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the usual ℓ_1 and ℓ_∞ norms.
Representing tournament structure: Estimating Θ requires specifying which tournament structures the model can represent. The Hodge decomposition separates Θ into a transitive component and a cyclic residual (Jiang et al., 2011; Balduzzi et al., 2018):

$$\Theta = \Theta^{\mathrm{trans}} + \Theta^{\mathrm{cyc}}, \qquad \Theta^{\mathrm{trans}}_{ij} = u_i - u_j, \quad u = \frac{1}{M}\Theta\mathbf{1}, \quad \Theta^{\mathrm{cyc}}\mathbf{1} = \mathbf{0}. \tag{1}$$

BT uses exactly the transitive parameterization Θ^{trans} (Bradley & Terry, 1952), thus cannot represent Θ^{cyc}. To represent both transitive and cyclic structure with adjustable capacity, we adopt the Disc-k model (Bertrand et al., 2023). It parameterizes antisymmetric matrices of rank at most $2k$ using $2kM$ parameters. Specifically, $\Theta_{ij} = c_i^\top \Omega_k c_j,\ c_i \in \mathbb{R}^{2k}$ and $\Omega_k = I_k \otimes \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

Estimation from finite votes: We estimate Θ from finite votes, using this parameterization to fit noisy observed comparisons and predict missing entries. Let $\mathcal{E} \subseteq \binom{[M]}{2}$ denote the observed pairs. For each $(i, j) \in \mathcal{E}$, we observe $N_{ij} \geq 1$ comparisons with W_{ij} wins for i, giving empirical win rate $\widehat{P}_{ij} = W_{ij}/N_{ij}$. The total sample size is $n = \sum_{\mathcal{E}} N_{ij}$. Pairs are sampled unevenly, so giving each vote equal weight amplifies frequently compared pairs. Under model misspecification, different sampling rules can therefore yield different rankings for the same tournament (Tang et al., 2025). We therefore assign each observed pair a fixed reference weight $q_{ij} > 0$, independent of its sampling frequency, with $\sum_{(i,j)\in\mathcal{E}} q_{ij} = 1$, and uniform weights give $q_{ij} = 1/|\mathcal{E}|$. Section 3 analyzes how model capacity and the available votes jointly determine matrix estimation error.

Desired ranking behavior. We now consider how rankings should reflect the estimated tournament. We introduce the axiomatic properties examined in our analysis; further social-choice background can be found in Fishburn (1984); Brandl et al. (2016); Lanctot et al. (2023)
Condorcet consistency: A model c that beats every rival, $A_{cj} > 0$ for all $j \neq c$, is a strict Condorcet winner. Condorcet consistency requires it to rank uniquely first whenever it exists.
Independence of irrelevant alternatives and clone consistency: Adding or deleting an irrelevant model or duplicating an existing one should preserve the relative order of the other models. We consider these changes with the pairwise win rates between the original models held fixed.

Beyond satisfying social-choice properties, a practical leaderboard must be robust to finite-sample estimation noise. The population margins $A^\star$ are unobserved, so ranking methods are applied to estimates $\widehat{A}$ that vary across samples. A reliable ranking method should exhibit statistical reproducibility: small changes in the estimated comparisons should not lead to substantially different conclusions. In Section4, we analyze this sensitivity to matrix estimation error.

However, retaining competitive structure in the estimated matrix does not ensure that the ranking responds to it. Next, we introduce four ranking methods on the same complete margin matrix A.

Fitting scalar scores: BT and SCO. Bradley-Terry (BT) model conflates estimation with ranking: its parameters are trained to both predict pairwise win rates and serve as ranking scores. As discussed after Eq.1, this parameterization represents only transitive logit structure and orders models by their average win rates (Siththaranjan et al., 2024), which need not place a Condorcet winner first.

Alternatively, Soft Condorcet Optimization (SCO) addresses BT's failure on Condorcet consistency through a sigmoid surrogate of the Kemeny–Young objective (Lanctot et al., 2025). At the global minimum of its nonconvex loss, SCO assigns the strict Condorcet winner a maximum score. But, this guarantee alone does not ensure independence from irrelevant alternatives or clone consistency.

Choosing an opponent distribution: MaxL and QRE. Probabilistic social choice provides rules with social-choice guarantees by selecting lotteries over models. Maximal lotteries (MaxL) select a strict Condorcet winner with probability one. They preserve the admissible lottery set under strict-loser insertion and, after aggregating clone probabilities, exact duplication (Brandl et al., 2016).

MaxL achieves these properties by choosing a mixture that no model beats on average, even with non-transitive comparisons (Fishburn, 1984). In the corresponding symmetric two-player zero-sum game, each player selects a model, with payoffs A_{ij} and $-A_{ij}$. Against a lottery $p \in \Delta_M := \{p \geq 0 : \mathbf{1}^\top p = 1\}$, model i has margin $(Ap)_i$, so the MaxL are $\mathcal{M}(A) = \{p \in \Delta_M : Ap \leq 0\}$.

A selected maximal lottery can still change substantially under sampling noise. Quantal response equilibrium (QRE) adds entropy to the same game to control this sensitivity (McKelvey & Palfrey, 1995; Sokota et al., 2023). For player lotteries $x, y \in \Delta_M$, temperature $\tau > 0$, and entropy $H(p) = -\sum_i p_i \log p_i$, the regularized game and its equilibrium conditions are

$$\begin{aligned} &\max_{x\in\Delta_M} \min_{y\in\Delta_M} \left\{x^\top Ay + \tau H(x) - \tau H(y)\right\}, \\ &p^\tau = \mathrm{softmax}(Ap^\tau/\tau), \qquad s^\tau = Ap^\tau. \end{aligned} \tag{2}$$

Here $\mathrm{softmax}(z)_i = e^{z_i}/\sum_j e^{z_j}$. The unique equilibrium can be computed efficiently and has both players use p^τ. The scores s^τ are expected margins against this shared opponent pool. At fixed positive temperature, the lottery and scores vary continuously with A, while exact MaxL guarantees may be relaxed.
Following this structure, Section 3 first bounds the error in the estimated comparisons; Section 4 then studies how temperature governs competitive guarantees and sensitivity to this error.

3 ESTIMATING THE MATRIX

Practical leaderboards often leave some model pairs unobserved, whereas the MaxL and QRE formulations above require a complete margin matrix. We therefore study how to estimate the full tournament from the observed pairwise comparisons using a regularized DISC-k model. We separate approximation, regularization, and sampling error and derive a finite-sample bound on the error of the fitted matrix. Full proofs are in Appendix B.

Estimator. We estimate $\Theta^\star$ by minimizing the following logistic loss with reference weight q and a nuclear-norm penalty over the DISC-k class $\mathbb{D}_k(\alpha) = \{\Theta = C\Omega_k C^\top : \|\Theta\|_{\max} \le \alpha\}$:

$$\widehat{\Theta}_k \in \operatorname*{arg\,min}_{\Theta \in \mathbb{D}_k(\alpha)} \Big\{ \widehat{\mathcal{L}}(\Theta) + \tfrac{\lambda}{2}\|\Theta\|_* \Big\}, \quad \widehat{\mathcal{L}}(\Theta) = \sum_{(i,j)\in\mathcal{E}} q_{ij}\big[\psi(\Theta_{ij}) - \widehat{p}_{ij}\Theta_{ij}\big], \tag{3}$$

where $\psi(t) = \log(1 + e^t)$, $\lambda \ge 0$, and $\|\cdot\|_*$ denotes the nuclear norm.

Sources of estimation error. We compare the empirical fit $\widehat{\Theta}_k$ with two population counterparts to distinguish the effects of limited capacity, regularization, and finite sampling. Replacing $\widehat{p}_{ij}$ with $P^\star_{ij}$ in equation 3 gives a regularized population fit $\Theta^{\dagger,\lambda}_k$. Removing the penalty gives an unregularized population fit $\Theta^\dagger_k$ within the same class $\mathbb{D}_k(\alpha)$. The estimation error then decomposes as

$$\widehat{\Theta}_k - \Theta^\star = \underbrace{(\Theta^\dagger_k - \Theta^\star)}_{\text{(I) Approximation}} + \underbrace{(\Theta^{\dagger,\lambda}_k - \Theta^\dagger_k)}_{\text{(II) Regularization}} + \underbrace{(\widehat{\Theta}_k - \Theta^{\dagger,\lambda}_k)}_{\text{(III) Sampling}}. \tag{4}$$

Approximation error arises from fitting the population comparisons within a restricted model class. Regularization error records the shift in this population fit caused by the nuclear-norm regularization. Win rates estimated from finite votes fluctuate around their population values, leading to sampling error in the empirical fit relative to the regularized population fit. Additional votes affect only the sampling term in this decomposition. We next derive the finite-sample error bound to examine the trade-off between capturing tournament structure and controlling sampling error.

Conditions for the error bound. The empirical loss in equation 3 uses only observed pairs, while our goal is to control error over the full matrix. We assume **(A1) Boundedness:** $\|\Theta^\star\|_{\max} \le \alpha < \infty$. This gives the logistic loss a positive curvature floor $\gamma_\alpha = \sigma'(\alpha) > 0$ over the feasible range. **(A2) Restricted strong convexity (RSC).** The empirical loss cannot distinguish fits that agree on observed pairs but differ on missing ones. We therefore need a condition under which small differences on observed pairs imply small differences across the full matrix. We assume RSC at the population fit $\Theta^\dagger_k$, with constant $\nu > 0$ (Negahban & Wainwright, 2012). This constant measures how strongly the observed pairs constrain the full matrix; smaller values imply weaker constraint. The precise condition is stated in Appendix B, Definition 1.

Finite-sample guarantee. Fix $\mathcal{E}$, N_{ij}, and q_{ij}, and assume mutually independent Bernoulli votes with means $P^\star_{ij}$. For a confidence level $1 - \delta$, with $\delta \in (0, 1)$, define the sampling term $\zeta_\delta = \sqrt{2d_\star \log(2M/\delta)} + \frac{2}{3}q_\star \log(2M/\delta)$, where $d_\star = \max_i \sum_{j:\{i,j\}\in\mathcal{E}} q^2_{ij}\sigma'(\Theta^\star_{ij})/N_{ij}$ and $q_\star = \max_{(i,j)\in\mathcal{E}} q_{ij}/N_{ij}$. $\zeta_\delta \to 0$ as every observed-pair count grows. The underlying concentration bound is given in Appendix B, Proposition 5.

Theorem 1 (Error of the DISC-k fit)**.** *Under* (A1)–(A2) *and the sampling assumptions stated above, let $\widehat{\Theta}_k$ be a global minimizer of equation 3. Write $B_k = \|\Theta^\dagger_k - \Theta^\star\|_F$ for the approximation error and set $q_{\max} = \max_{(i,j)\in\mathcal{E}} q_{ij}$. For $\delta \in (0, 1)$, if $\lambda \ge 2\zeta_\delta + (q_{\max}/2)B_k$, with ζ_δ defined above, then with probability at least $1 - \delta$, $\|\widehat{\Theta}_k - \Theta^\star\|_F \le B_k + 6\sqrt{k}\lambda/(\gamma_\alpha\nu)$. In particular, choosing $\lambda = 2\zeta_\delta + (q_{\max}/2)B_k$ gives*

$$\|\widehat{\Theta}_k - \Theta^\star\|_F \le \left(1 + \frac{3q_{\max}\sqrt{k}}{\gamma_\alpha\nu}\right) B_k + \frac{12\sqrt{k}}{\gamma_\alpha\nu}\zeta_\delta =: R_\Theta(\delta). \tag{5}$$

In equation 5, regularization is included in both terms through the chosen λ. Increasing all observed-pair counts reduces the sampling term but leaves B_k unchanged at fixed k, $\mathcal{E}$, and q. A larger k expands the representable class which may reduce the structural bias B_k, but increases the contribution of sampling error through the $\sqrt{k}$ factor. In section 5.3, we provide empirical evidence for this trade-off through a recovery experiment that varies model capacity and the comparison budget.

4 RANKING FROM ESTIMATIONS

Section 3 quantified the error in estimating a tournament from finite votes. How this uncertainty affects the leaderboard depends on the ranking method applied to the estimated matrix. Table 1 compares the social-choice guarantees and ranking properties of the methods introduced in Section 2. In this section, we quantify two effects of QRE's temperature: how closely QRE satisfies social-choice properties and how much its lotteries and payoff scores change under estimation error.

Table 1: Comparison of ranking readouts on a common complete margin matrix, with equal-edge BT/SCO fits. ✓: the stated guarantee; ×: no general guarantee; ε: approximate guarantee.

Ranking Methods	Condorcet Consistency	IIA/Clone Consistency	Total Ordering	Continuity of Scores/Lotteries
BT	×	×	✓	✓
SCO	✓ (global optimum)	×	✓	×
MaxL	✓	✓	×	×
QRE$_\tau$	✓ ($\tau < \tau_{\mathrm{CW}}$)	$\varepsilon_{\mathrm{IIA}}$	✓	✓

Temperature-dependent social-choice guarantees At positive temperature, QRE assigns nonzero probability to every model. A strict Condorcet winner therefore shares probability with its rivals, and a new entrant can have bounded affect on existing scores even when the original pairwise win rates remain unchanged. Theorem 2 gives a temperature condition under which the Condorcet winner still ranks first and bounds how much an arbitrary entrant can change the original models' score differences. Define sensitivity factor as $\beta(A) := \sup_{\mathbf{1}^\top x=0,\, \|x\|_1=1} \|Ax\|_\infty \le 1$.

Theorem 2 (Social-choice guarantees of QRE)**.** *Fix $\tau > 0$ and write $s = s^\tau(A)$ as in Eq. 2.* (i) Approximate Condorcet consistency. *If c is a strict Condorcet winner, let $g_{\mathrm{CW}} = \min_{j \neq c} A_{cj} > 0$ be its smallest winning marging and $a_{\max} = \|A\|_{\max}$ be the largest margin across the entire matrix. Then QRE ranks c strictly first whenever*

$$\tau < \tau_{\mathrm{CW}} := g_{\mathrm{CW}}^2/(a_{\max} \log M).$$

(ii) Approximate independence of irrelevant alternatives. *Let $A' \in [-1,1]^{(M+1)\times(M+1)}$ be antisymmetric after inserting u, with $A'_{ij} = A_{ij}$ for $i, j \in [M]$. Write $a_i = A'_{iu}$ for model i's margin against u. Let $s' = s^\tau(A')$ and $\mu = p^\tau_u(A')$ be the new scores and entrant probability. For all $i, j \in [M]$,*

$$\left|(s'_i - s'_j) - (s_i - s_j)\right| \le \mu \left(\left|(a_i - a_j) - (s_i - s_j)\right| + \frac{2\beta(A)}{\tau} \|a - s\|_\infty \right) =: \varepsilon^{ij}_{\mathrm{IIA}}.$$

Let $\Gamma := \max_{k\in[M]} s'_k - s'_u$ denote the entrant's score deficit relative to the candidate with highest score, when $\Gamma \ge 0$, we have $\mu \le (1 + e^{\Gamma/\tau})^{-1} \le e^{-\Gamma/\tau}$.

These bounds show how low temperature can preserve social-choice guarantees as QRE approaches MaxL. QRE ranks a strict Condorcet winner first for $\tau < \tau_{\mathrm{CW}}$, and the IIA bound vanishes exponentially fast as $\tau \downarrow 0$ for any positive Γ, a condition satisfied by any strict loser. We generalize this approximate IIA result to approximate clone consistency in Appendix C.5. The next subsection quantifies the statistical cost: lower temperature weakens the stability bound under estimation error.

Stability under estimation error Even when the underlying pairwise win rates remain fixed, finite votes yield estimates that vary across samples. A selected MaxL lottery can change discontinuously under arbitrarily small changes in the estimated win rates. Figure 1(c) illustrates the consequence: resampling the same votes can substantially change MaxL's output distribution. At positive temperature, QRE's lotteries and scores vary continuously with the matrix. Using the entropy-regularized game framework (Sokota et al., 2023), we bound their sensitivity as a function of temperature

Theorem 3 (Stability of QRE ranking)**.** *For antisymmetric $A, B \in [-1,1]^{M\times M}$, fixed $\tau > 0$, and $\varepsilon = \|A - B\|_{\max}$, we have: $\|p^\tau(A) - p^\tau(B)\|_1 \le \frac{\varepsilon}{\tau}$, and $\|s^\tau(A) - s^\tau(B)\|_\infty \le (1 + \tau^{-1})\varepsilon$.*

The $1/\tau$ factor makes the statistical cost of low temperature explicit: smaller τ requires more accurate matrix estimates to maintain the same stability guarantee. Continuous scores can still exchange order near ties, so a reported lead must exceed the possible movement of both scores.

Combining the estimation and stability bounds gives a sufficient gap for a reported QRE lead to hold at the population level. Let $\widehat{s}$ be the QRE scores of the fitted tournament at fixed $\tau > 0$, and let $R_\Theta(\delta)$ be Theorem 1's error bound. Under that theorem's assumptions, with probability at least $1-\delta$, simultaneously for all pairs, $\widehat{s}_i - \widehat{s}_j > (1+\tau^{-1})R_\Theta(\delta) \Longrightarrow s_i^\tau(A^\star) > s_j^\tau(A^\star)$. Appendix C.4 gives the derivation.

Temperature therefore changes both QRE's social-choice guarantees and its sensitivity to sampling noise. As $\tau \downarrow 0$, QRE approaches the MaxL, while the stability bound weakens. As $\tau \to \infty$, its scores approach $A\mathbf{1}/M$, the average win rate, and the stability bound tightens. The guarantees in Theorem 2 and the perturbation bounds above quantify these two effects. Section 5 examines their empirical consequences on Chatbot Arena and Image Bench through bootstrap resampling, controlled changes to the candidate pool and comparisons, and matrix completion.

5 Experiments

5.1 What do the observed comparisons support?

We use two complete pairwise comparison matrices: one from a subset of Chatbot Arena (28 models, 71,866 votes) (Chiang et al., 2024), and the other from Datapoint Image Bench (30 models, 2,161,160 annotations on 500 prompts across ten categories) (Datapoint AI, 2026). We use 2,000 matched bootstrap resamples, drawing votes independently for Arena and prompts for Image.

We summarize score separation using *confidence-interval-separated pairs* and the *leader rank spread* similar to (Datapoint AI, 2026). A pair is separated when the two models' 95% score confidence intervals do not overlap; a leader rank spread of $[1, 2]$ indicates that one rival remains unseparated from the full-data leader. Appendix D.1 provides the detailed definitions and experimental setup. We report Kendall–tau distance (Lanctot et al., 2025) as the mean fraction of model pairs whose ordering reverses in bootstrap resamples relative to each method's full-data result.

Table 2: Score separation and sampling stability on 2,000 matched resamples. †SCO results are solver-conditional due to its nonconvex loss.

	Arena (28 models)				Image (30 models)			
Ranking	Separated pairs	Leader rank spread	Kendall distance	Leader change	Separated pairs	Leader rank spread	Kendall distance	Leader change
BT/Elo	314/378	$[1, 1]$	2.68%	0.00%	374/435	$[1, 2]$	1.85%	8.40%
SCO†	312/378	$[1, 1]$	2.72%	3.90%	373/435	$[1, 2]$	2.11%	7.50%
QRE ($\tau = .01$)	107/378	$[1, 4]$	10.83%	8.05%	279/435	$[1, 2]$	4.81%	5.60%
QRE ($\tau = .075$)	247/378	$[1, 1]$	4.97%	0.05%	354/435	$[1, 2]$	1.83%	2.45%
QRE ($\tau = .5$)	312/378	$[1, 1]$	2.51%	0.00%	374/435	$[1, 2]$	1.80%	6.60%
ML support	—	—	—	27.45%	—	—	—	6.25%

The same comparisons yield different amounts of score separation. In Arena, BT/Elo and SCO separate 314 and 312 of 378 pairs, compared with 247 for QRE at $\tau = .075$. BT fits scores and accounts for average performance of each candidates. SCO uses the full pairwise soft-disagreement objective to fit a scalar ordering. QRE and MaxL directly read the empirical matrix without fitting any estimation model, with QRE adding entropy regularization. These choices combine different structural constraints with different sensitivities to estimation noise. The resulting separation concerns each method's own score representation; more separated pairs alone do not establish greater fidelity to the tournament structure. All scalar rankings leave one rival unseparated from Image's leader.

Sampling stability depends on the strength of regularization. At $\tau = .075$, QRE changes its leader in 0.05% of Arena resamples and 2.45% of Image resamples. MaxL's support changes in 27.45% and 6.25%, respectively. Increasing temperature reduces pair reversal rates under resampling. In Arena, Kendall–tau distance decreases from 10.83% to 4.97% to 2.51% at $\tau = .01, .075, .5$, respectively. On Image, increasing τ from .075 to .5 barely changes Kendall distance (1.83% to 1.80%), while leader changes increase from 2.45% to 6.60%.

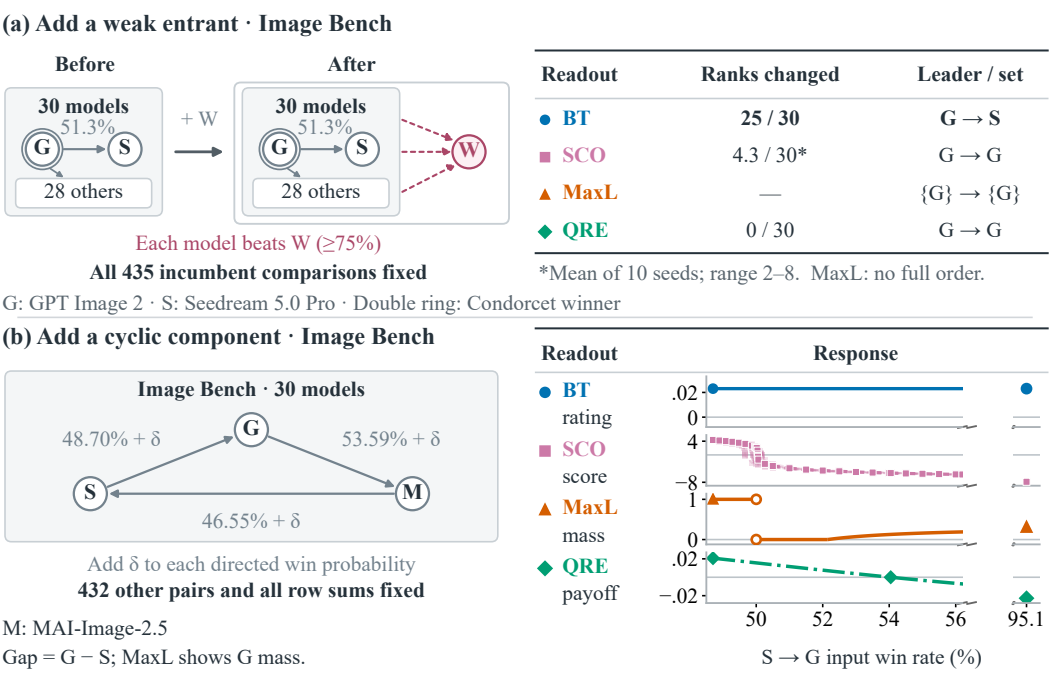


Figure 2: **Fixed matchups need not fix BT; changed matchups need not change BT.** (a) W is chosen to disrupt BT; counts measure changed ranks (SCO: evaluated on ten seeds). (b)BT, SCO, and QRE show GPT-minus-Seedream gaps on separate scales; MaxL shows GPT mass. Open circles mark one-sided limits at 50%; isolated points mark the 95.11% endpoint. SCO connects evaluated inputs from ten paired seeds. QRE uses $\tau = .075$.

5.2 How do ranking respond to controlled changes?

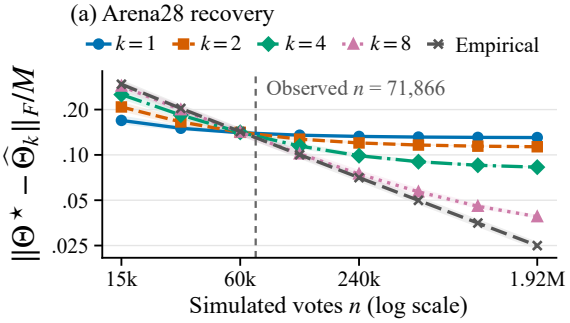
BT can change ranks when original comparisons stay fixed, yet remain unchanged when comparisons change along a cycle. We compare these responses across rankings on empirical tournaments. Figure 2a shows that adding a weak entrant changes BT's leader while GPT Image 2 remains the Condorcet winner. Arena provides a deletion example: removing two successive strict losers changes ten of 26 survivor rank positions under BT and swaps ChatGPT-4o and o3 in second and third place. In both cases, QRE at $\tau = .075$ preserves the common-model ordering, MaxL preserves its singleton support, and SCO retains the leader across ten paired seeds (Appendix D.3).
This example illustrates the tradeoff between sampling stability and social-choice guarantees in QRE. Higher temperature spreads QRE probability more broadly, giving the weak entrant more mass in this example. Theorem 2(ii) bounds the resulting changes in score gaps in terms of this mass. This insertion changes $0, 0, 4$ incumbent ranks at $\tau = .01, .075, .5$, respectively.

Together with Table 2, this shows that stronger smoothing can improve whole-ranking sampling stability at the cost of robustness to irrelevant alternatives.

Figure 2b gives the complementary result: cyclic changes leave BT unchanged because it is incapable of representing the cyclic components. Even when Seedream's direct win rate against GPT reaches 95.11%, BT still predicts 49.42%. MaxL switches support at the direct 50% boundary, whereas QRE at $\tau = .075$ changes its score gap continuously (Appendix D.2).

5.3 How does completion change the reported answer?

Theorem 1 decomposes estimation error into approximation, regularization, and sampling terms. To illustrate the interaction between approximation and sampling, we set $\lambda = 0$, freeze the empirical Arena28 tournament as the population, and draw 2,000 matched vote samples at different budget n. Figure 3a reports $\|\Theta^\star - \widehat{\Theta}_k\|_F/M$. At small budgets, smaller k gives lower recovery error, consistent with reduced sensitivity to sampling noise. As more votes reduce sampling error, the approximation



(b) Arena48 prediction

Estimator	NLL ↓	Cyclic fraction (%)
BT	**0.66547**	0.00
DISC-1	0.66596	1.55
DISC-2	0.66731	3.35
DISC-4	0.66771	4.59
DISC-8	0.66742	4.82

Figure 3: **Capacity affects recovery and fitted cyclic structure.** (a) Arena28 logit recovery from a frozen empirical population ($\lambda = 0$); lines and bands show medians and 5th–95th sampling percentiles over 2,000 resamples per budget. (b) Arena48 held-out NLL and cyclic fractions.

Table 3: **Leaderboard outcomes depend on matrix construction and ranking.** Arena48 point top sets with exact recurrence (%) in parentheses over 2,000 bootstrap resamples. MaxL reports its first tier; QRE uses $\tau = 0.075$. G: Gemini 2.5 Pro; P: its preview.

Ranking	Laplace	BT full	DISC-1 full	DISC-2 full	DISC-4 full
QRE	$\{G\}$ (99.80)	$\{G\}$ (99.95)	$\{G\}$ (96.25)	$\{G\}$ (97.90)	$\{G\}$ (98.60)
MaxL	$\{G, P\}$ (76.50)	$\{G\}$ (99.95)	$\{G\}$ (92.45)	$\{G\}$ (92.25)	$\{G\}$ (90.10)
Observed-pair fitting:	BT $\{G\}$ (99.95)		SCO $\{G, P\}$ (38.35)		

limits of low-capacity fits become apparent: their errors plateau, while larger k achieves lower error and the empirical estimator eventually performs best. The curves thus illustrate why additional data and additional capacity address different sources of error. We also evaluate held-out prediction on Arena48, formed by retaining all models with more than 1,000 votes from the complete 53-model Chatbot Arena pool. DISC-k attains held-out NLL close to BT's while allowing nonadditive structure: its fitted logits contain cyclic components that BT excludes by construction (Figure 3b).

We next combine estimation with ranking on Arena48. Table 3 compares direct ranking with ranking after matrix construction. BT and SCO fit the observed pairs directly but report different top sets. For QRE and MaxL, which require a complete matrix, we compare fitted reconstruction with Laplace smoothing, a standard correction used in MaxL evaluation (Khalaf et al., 2026). Laplace smoothing shrinks observed win rates toward $1/2$ and assigns every unobserved pair exactly $1/2$, without using evidence from either model's other comparisons. Thus, treating missing pairs as ties is a strong structural assumption. BT and DISC instead share information across pairs through a fitted model, reconstructing both observed and missing entries. At the point estimates, this distinction changes MaxL's first tier from G, P under Laplace smoothing to G under fitted reconstruction, while QRE ranks G first across all constructions. Bootstrap recurrence also varies across constructions. These frequencies describe reproducibility under each pipeline's assumptions. Appendix D.6 examines observed-entry reconstruction and additional temperatures.

6 Conclusion

Evaluating AI models remains challenging without an objective ground truth. In this work, we theoretically link tournament structure, model capacity, and sample size to matrix estimation errors and QRE rankings. Empirically, we demonstrate that the way observed comparisons are reconstructed and evaluated significantly impacts reproducibility. This requires careful reporting of the chosen estimator, the ranking method, and the uncertainty of the outcome to ensure interpretable results. A primary limitation of our QRE framework is its dependence on the temperature parameter, which alters both the final ranking and its sensitivity. Additionally, while our approach provides strong stability guarantees, it does not currently recover the Smith set structure (Smith, 1973), leaving this theoretical extension for future work.

STATEMENT ON THE USE OF AI.

Generative AI tools (large language models and LLM-based coding agents) were used in this work in the role of research assistants. We used them to cross-check proofs, evaluate the soundness of technical arguments, and provide critical feedback on the precision of our claims. We did not use generative AI to formulate the mathematical claims, design the methodology or experiments, generate or clean data, or interpret the results. Additionally, we used generative AI to survey the related literature, draft and edit parts of the manuscript for clarity, and, through coding agents, execute our experiment scripts, run hyperparameter grid searches, and edit run configurations and parameters. Implementation is assisted but not designed by AI. We reviewed all AI-assisted work, and every number reported in the text was verified against the archived results. We take responsibility for the final content of this work, including the text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses two publicly released preference datasets (LMArenas arena-human-preference-140k and Datapoints Image Bench), both licensed under CC BY 4.0, and only their aggregated pairwise vote counts. We do not use redistribute, or analyze conversation text, prompts, images, or annotator-level information, and we made no attempt to re-identify annotators. No new human-subject data were collected. Our interventions (removing models, adding a synthetic entrant, and shifting a small cycle of win rates) are counterfactual analyses of frozen historical snapshots. No live leaderboard was affected, no Arena votes were fabricated, and none of the perturbation paths is presented as an executable attack. Reported rank changes for named models are properties of the ranking rules under these interventions, not assessments of current model quality. The authors have no affiliation with the dataset providers or the model developers involved.

REPRODUCIBILITY STATEMENT

The appendices provide proofs, data-processing details, and experimental protocols. Both datasets are publicly available. Anonymous supplementary code includes the method implementations, experiment routines, and parameter settings.

REFERENCES

David Balduzzi, Karl Tuyls, Julien Pérolat, and Thore Graepel. Re-evaluating evaluation. In *Advances in Neural Information Processing Systems 31 (NeurIPS)*, pp. 3272–3283, 2018.

Quentin Bertrand, Wojciech Marian Czarnecki, and Gauthier Gidel. On the limitations of the Elo: Real-world games are transitive, not additive. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 206, pp. 2905–2921, 2023.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

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.

Florian Brandl, Felix Brandt, and Hans Georg Seedig. Consistent probabilistic social choice. *Econometrica*, 84(5):1839–1880, 2016.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Michael Jordan, Joseph E Gonzalez, et al. Chatbot arena: An open platform for evaluating llms by human preference. In *Forty-first International Conference on Machine Learning*, 2024.

Datapoint AI. Text-to-image human preferences 2m: the datapoint image bench voting record, 2026. URL https://huggingface.co/datasets/datapointai/text-2-image-human-preferences-2m. 2,161,160 leaderboard-grade pairwise votes over 30 models, 500 prompts, full round-robin.

Mark A. Davenport, Yaniv Plan, Ewout van den Berg, and Mary Wootters. 1-bit matrix completion. *Information and Inference*, 3(3):189–223, 2014.

Peter C. Fishburn. Probabilistic social choice based on simple voting comparisons. *The Review of Economic Studies*, 51(4):683–692, 1984.

Jenny Y. Huang, Yunyi Shen, Dennis Wei, and Tamara Broderick. Dropping just a handful of preferences can change top large language model rankings. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=jNiEMDsRgc.

Yangsibo Huang, Milad Nasr, Anastasios Nikolas Angelopoulos, Nicholas Carlini, Wei-Lin Chiang, Christopher A. Choquette-Choo, Daphne Ippolito, Matthew Jagielski, Katherine Lee, Ken Liu, Ion Stoica, Florian Tramèr, and Chiyuan Zhang. Exploring and mitigating adversarial manipulation of voting-based leaderboards. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 25654–25671. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/huang25z.html.

Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial hodge theory. *Mathematical Programming*, 127(1):203–244, 2011.

Hadi Khalaf, Serena L. Wang, Daniel Halpern, Itai Shapira, Flavio P. Calmon, and Ariel D. Procaccia. Robust AI evaluation through maximal lotteries. *arXiv preprint arXiv:2602.21297*, 2026.

Marc Lanctot, Kate Larson, Yoram Bachrach, Luke Marris, Zun Li, Avishkar Bhoopchand, Thomas Anthony, Brian Tanner, and Anna Koop. Evaluating agents using social choice theory. *arXiv preprint arXiv:2312.03121*, 2023.

Marc Lanctot, Kate Larson, Michael Kaisers, Quentin Berthet, Ian Gemp, Manfred Diaz, Roberto-Rafael Maura-Rivero, Yoram Bachrach, Anna Koop, and Doina Precup. Soft Condorcet optimization for ranking of general agents. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 1253–1262, 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.

Marie Jean Antoine marquis de Condorcet. *Essai sur l'application de l'analyse a la probabilite des decisions: rendues a la pluralite de voix*. De l'Imprimerie royale, 1785.

Richard D. McKelvey and Thomas R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995.

Rui Min, Tianyu Pang, Chao Du, Qian Liu, Minhao Cheng, and Min Lin. Improving your model ranking on chatbot arena by vote rigging. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=5cDc71jLc1.

Rémi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, Andrea Michi, Marco Selvi, Sertan Girgin, Nikola Momchev, Olivier Bachem, Daniel J. Mankowitz, Doina Precup, and Bilal Piot. Nash learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235, pp. 36743–36768, 2024.

Sahand Negahban and Martin J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, 13:1665–1697, 2012.

Hosna Oyarhoseini, Jimmy Lin, and Amir-Hossein Karimi. A unified perturbation framework for analyzing leaderboard stability and manipulation. *arXiv preprint arXiv:2605.15761*, 2026.

Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell. Distributional preference learning: Understanding and accounting for hidden context in rlhf. In *International Conference on Learning Representations*, volume 2024, pp. 27448–27471, 2024.

John H. Smith. Aggregation of preferences with variable electorate. *Econometrica*, 41(6): 1027–1041, 1973. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1914033.

Samuel Sokota, Ryan D'Orazio, J. Zico Kolter, Nicolas Loizou, Marc Lanctot, Ioannis Mitliagkas, Noam Brown, and Christian Kroer. A unified approach to reinforcement learning, quantal response equilibria, and two-player zero-sum games. In *International Conference on Learning Representations (ICLR)*, 2023.

Shange Tang, Yuanhao Wang, and Chi Jin. Is elo rating reliable? a study under model misspecification. *arXiv preprint arXiv:2502.10985*, 2025.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8 (12):1230, May 2015. ISSN 1935-8237. doi: 10.1561/2200000048. URL https://doi.org/10.1561/2200000048.

A Notation and Formal Readout Definitions

Table 4: Notation used throughout. Vector ℓ_∞ and matrix maximum-entry norms are distinguished explicitly.

Symbol	Meaning
M, k, n	Number of models, DISC capacity, total votes
$P^\star$, $\Theta^\star$, $A^\star$	Population preference credits, logits, margins
$\widehat{\Theta}$, $\widehat{A}$	Fitted logits and induced margins
$\mathcal{E}$, N_{ij}, q_{ij}	Observed pairs, pair counts, fixed reference weights
α, λ	Logit envelope and nuclear-norm penalty
$\Theta_k^\dagger$	Population rank-constrained likelihood target
b_k, $\mathcal{B}_k = Mb_k$	RMS approximation error and total Frobenius bias
$\widetilde{\Theta}_k$, $\mathcal{T}_{k-1}$	Transitive-plus-cyclic comparator and discarded cyclic tail
$p^\tau(A)$, $s^\tau(A)$	QRE opponent distribution and margin scores
Δ_M	Probability simplex on the M models
$b = A\mathbf{1}/M$	Average margin against the uniform opponent pool
$a_{\max}$, g_{CW}	Largest absolute margin, strict CW margin
$\varepsilon_{\mathrm{CW}}$, $\varepsilon_{\mathrm{ins}}$	Bounds on missing CW mass and incumbent gap change
τ, η, B	QRE temperature, SCO smoothing, SCO score bound
$\|Z\|_{\max}$, $\|Z\|_F/M$	Maximum-entry norm and entrywise RMS norm

B Estimation: Statements and Proofs

Estimation from finite votes (from Section 2: We estimate Θ from finite votes, using this parameterization to fit noisy observed comparisons and predict missing entries. Let $\mathcal{E} \subseteq \binom{[M]}{2}$ denote the observed pairs. For each $(i,j) \in \mathcal{E}$, we observe $N_{ij} \geq 1$ comparisons with W_{ij} wins for i, giving empirical win rate $\widehat{P}_{ij} = W_{ij}/N_{ij}$. The total sample size is $n = \sum_{\mathcal{E}} N_{ij}$. Pairs are sampled unevenly, so giving each vote equal weight makes frequently compared pairs more influential in estimation. Under model misspecification, different sampling rules can therefore yield different rankings for the same tournament (Tang et al., 2025). We therefore assign each observed pair a fixed reference weight $q_{ij} > 0$, independent of its sampling frequency, with $\sum_{(i,j)\in\mathcal{E}} q_{ij} = 1$, and uniform weights give $q_{ij} = 1/|\mathcal{E}|$. Sampling frequencies then only affect the variance of empirical win rates without changing the population fitting target. Section 3 analyzes how model capacity and the available votes jointly determine matrix estimation error.

B.1 PROOF SKETCH OF THE MAIN THEOREM

We restate Theorem 1 from Section 3 and outline its proof.

Theorem 1 (restated). *Assume* (A1)–(A2)*, let* $\widehat{\Theta}_k$ *be a global minimizer of equation 3, and fix* $\delta \in (0,1)$*. Write* $q_{\max} = \max_{\mathcal{E}} q_{ij}$ *and let* $B_k = \|\Theta_k^\dagger - \Theta^\star\|_F$ *be the irreducible approximation bias of the disc-*k *class. If the regularization satisfies* $\lambda \geq 2\zeta_\delta + \frac{q_{\max}}{2} B_k$*, with* ζ_δ *as in equation 6, then with probability at least* $1-\delta$*,*

$$\|\widehat{\Theta}_k - \Theta^\star\|_F \;\leq\; B_k + \frac{6\sqrt{k}\,\lambda}{\gamma_\alpha \nu}.$$

In particular, setting λ *at the equality gives*

$$\|\widehat{\Theta}_k - \Theta^\star\|_F \;\leq\; \Big(1 + \frac{3q_{\max}\sqrt{k}}{\gamma_\alpha \nu}\Big) B_k + \frac{12\sqrt{k}}{\gamma_\alpha \nu}\,\zeta_\delta.$$

We define

$$\zeta_\delta = \sqrt{2d_\star \log(2M/\delta)} + \tfrac{2}{3} q_\star \log(2M/\delta) \tag{6}$$

Two previously stated conditions/definitions about the setup are used in the proof: (A3) the observation model has fixed counts N_{ij}, fixed weights q_{ij} and independent outcomes, and (A4), that the optimizer returns a global minimizer (Appendix B.2 and Remark 1).

Step 1: Redefining the loss functions. Summing the cross-entropy over the N_{ij} votes of each pair and weighting by q_{ij} (derivation in Appendix B.3) gives the empirical loss, which is the loss in equation 3. Replacing the observed win rates by the true ones gives the population loss:

$$\widehat{\mathcal{L}}(\Theta) = \sum_{(i,j)\in\mathcal{E}} q_{ij}\big[\psi(\Theta_{ij}) - \widehat{P}_{ij}\Theta_{ij}\big], \qquad \overline{\mathcal{L}}(\Theta) = \sum_{(i,j)\in\mathcal{E}} q_{ij}\big[\psi(\Theta_{ij}) - P^\star_{ij}\Theta_{ij}\big],$$

with $\widehat{P}_{ij} = W_{ij}/N_{ij}$ and $P^\star_{ij} = \sigma(\Theta^\star_{ij})$. By Proposition 1, training the capability matrix C with weight decay solves the same problem as the nuclear-norm program in Θ, so the whole analysis works directly with Θ.

The population loss has no sampling noise. Minimizing it over the same class, without and with the penalty, gives two reference matrices:

$$\Theta_k^\dagger \in \operatorname*{arg\,min}_{\Theta\in\mathbb{D}_k(\alpha)} \overline{\mathcal{L}}(\Theta), \qquad \Theta_k^{\dagger,\lambda} \in \operatorname*{arg\,min}_{\Theta\in\mathbb{D}_k(\alpha)} \Big\{\overline{\mathcal{L}}(\Theta) + \tfrac{\lambda}{2}\|\Theta\|_*\Big\}.$$

$\Theta_k^\dagger$ is the best fit the disc-k class could reach with infinitely many votes, and $\Theta_k^{\dagger,\lambda}$ is the same fit when the penalty is also applied. By Lemma 8, $\Theta^\star$ itself minimizes $\overline{\mathcal{L}}$, so when $\operatorname{rank}\Theta^\star \leq 2k$ we can take $\Theta_k^\dagger = \Theta^\star$.

Step 2: split the error into three terms. As in equation 4 of the main paper, the total error is a sum of three terms:

$$\widehat{\Theta}_k - \Theta^\star = \underbrace{(\Theta_k^\dagger - \Theta^\star)}_{\textbf{(I) Approximation}} + \underbrace{(\Theta_k^{\dagger,\lambda} - \Theta_k^\dagger)}_{\textbf{(II) Regularization}} + \underbrace{(\widehat{\Theta}_k - \Theta_k^{\dagger,\lambda})}_{\textbf{(III) Estimation Variance}}.$$

Term (I) depends only on the capacity k, term (II) only on the penalty λ and neither is random. Term (III) is the only random term, and it comes from having finitely many votes.

The proof bounds (I) on its own, but bounds (II) and (III) *together*. Their sum is $\widehat{\Theta}_k - \Theta_k^\dagger$, so by the triangle inequality

$$\|\widehat{\Theta}_k - \Theta^\star\|_F \;\leq\; \underbrace{\|\Theta_k^\dagger - \Theta^\star\|_F}_{=B_k\ \text{(I)}} + \underbrace{\|\widehat{\Theta}_k - \Theta_k^\dagger\|_F}_{\text{(II)}+\text{(III)}}.$$

Therefore, it suffices to bound B_k (Step 4) and the distance from $\widehat{\Theta}_k$ to $\Theta_k^\dagger$ (Steps 5–7).

Step 3: tools/lemmas used for every term (Appendix B.4.1). Three facts about the loss are used throughout.

(a) *Loss identity* (Lemma 5). For any reference matrix Θ°, the change in loss from Θ° to Θ is a nonnegative curvature term plus a term that is exactly linear in $\Theta - \Theta^\circ$, with coefficient the gradient S at Θ° (see equation 12).

(b) *Curvature* (Lemma 6). When all entries lie in $[-\alpha, \alpha]$ (Assumption A1), the curvature term is at least $\gamma_\alpha/2$ times the weighted squared error, by equation 13. This turns a gap in loss into a squared distance.

(c) *Noise/bias split of the gradient* (Lemma 7). The empirical gradient is the population gradient plus the statistical noise matrix Z(see equation 14). The population gradient is zero at $\Theta^\star$ and grows at most like $\frac{q_{\max}}{4}\|\Theta - \Theta^\star\|_F$ away from it(see equation 15).

Step 4: Term (I): approximation (Appendix B.4.2). Proposition 2 compares $\Theta_k^\dagger$ with the rank-$2k$ Youla truncation of $\Theta^\star$, using the curvature bounds of Step 3 on both sides (see equation 16). For a complete pair set this gives $B_k \le e^{\alpha/2}\sqrt{q_{\max}/q_{\min}}\,\varsigma_{>k}$, where $\varsigma_{>k}$ measures the part of $\Theta^\star$ that lies beyond its first k blocks. This term is not random and does not shrink with more votes and it is zero when $\operatorname{rank}\Theta^\star \le 2k$.

Step 5: the main/master lemma (Appendix B.4.3). Lemma 10 is the main deterministic Lemma applied on loss functions. It takes any loss of the form in Lemma 5, a reference matrix $\Theta^\circ \in \mathbb{D}_k(\alpha)$, and a global minimizer $\Theta^\bullet$ of that loss plus the penalty. If $\lambda \ge 2\|S\|_{\mathrm{op}}$, where S is the gradient at Θ°, then $\|\Theta^\bullet - \Theta^\circ\|_F \le 6\sqrt{k}\,\lambda/(\gamma_\alpha\nu)$, which is equation 22. The proof has four parts:

(i) $\Theta^\bullet$ is optimal, so its penalized loss is no larger than that of Θ°; the loss identity of Step 3 rewrites this as equation 23.

(ii) The linear term is bounded by $\|S\|_{\mathrm{op}}$ times the nuclear norm of the error (Lemma 12(i)), and $\lambda \ge 2\|S\|_{\mathrm{op}}$ makes the penalty larger than this term.

(iii) Splitting the error into a low-rank part and a remainder (Lemma 9) shows that the error lies in the cone of Definition 1. This uses only optimality.

(iv) On that cone, (A2) in the form equation 21 and the curvature bound give the final inequality.

The lemma is used twice, always with reference matrix $\Theta_k^\dagger$:

Application	Loss	Minimizer $\Theta^\bullet$	Error controlled	Result
Regularization	$\overline{\mathcal{L}}$	$\Theta_k^{\dagger,\lambda}$	(II)	Proposition 3
Main theorem	$\widehat{\mathcal{L}}$	$\widehat{\Theta}_k$	(II) + (III)	Theorem 1

The first one gives the separate bound on (II) in Proposition 3. The second is Step 7.

Step 6: Statistical error from sampling noise (Appendix B.4.4). The only random object is the score matrix Z. By Proposition 4, its entries are independent, have mean zero, and have variance $q_{ij}^2\sigma'(\Theta_{ij}^\star)/N_{ij}$, which shrinks as a pair receives more votes. Proposition 5 writes Z as a sum over single comparisons and applies matrix Bernstein (Lemma 13): with probability at least $1-\delta$, $\|Z\|_{\mathrm{op}} \le \zeta_\delta$, where ζ_δ in equation 30 is the same as equation 6 in the main paper. We find that summing over single comparisons, rather than over pairs, is needed for the noise to vanish as $N_{ij} \to \infty$. Remark 4 explains what $d_\star$ says about how votes should be allocated.

Step 7: putting the pieces together (Appendix B.4.4). The proof of Theorem 1 applies the master lemma of Step 5 to the empirical loss $\widehat{\mathcal{L}}$, with reference $\Theta^\circ = \Theta_k^\dagger$ and minimizer $\Theta^\bullet = \widehat{\Theta}_k$.

(i) By Step 3(c), the gradient at $\Theta_k^\dagger$ is $S = Z + \nabla\overline{\mathcal{L}}(\Theta_k^\dagger)$, so $\|S\|_{\mathrm{op}} \le \|Z\|_{\mathrm{op}} + \frac{q_{\max}}{4}B_k$ by equation 15.

(ii) From Step 6, $\|Z\|_{\mathrm{op}} \le \zeta_\delta$. The condition $\lambda \ge 2\zeta_\delta + \frac{q_{\max}}{2}B_k$ in the theorem guarantees $\lambda \ge 2\|S\|_{\mathrm{op}}$.

(iii) The main/master lemma gives $\|\widehat{\Theta}_k - \Theta_k^\dagger\|_F \le 6\sqrt{k}\,\lambda/(\gamma_\alpha\nu)$, and the triangle inequality of Step 2 adds B_k. This is the first bound of the theorem.

(iv) Setting $\lambda = 2\zeta_\delta + \frac{q_{\max}}{2} B_k$ and collecting the common terms gives the second bound, equation 31.

The extra term $\frac{q_{\max}}{2} B_k$ in the condition on λ is the error due to misspecification: when the truth is outside the class, the population gradient at $\Theta_k^\dagger$ is not zero, and the regularization penalty must dominate it as well as the noise. The same bounds hold for the margin matrix after halving, using Lemma 1.

Step 8: rates and sample complexity (Appendix B.5.1). Corollary B.1 simplifies the main theorem to a balanced design: every pair is observed, each pair gets $\bar{N}$ votes, and all weights equal $\bar{q} = \binom{M}{2}^{-1}$. Then $\nu = q_{\max} = \bar{q}$ by Definition 1, and the weight $\bar{q}$ cancels from the bound. Writing $\bar{N}$ in terms of the total number of votes $n = \bar{N}\binom{M}{2}$ gives the per-entry rate equation 32. The proof uses $n \geq M \log(2M/\delta)$ so that the second term of ζ_δ is dominated by the first. The rate has two parts: the approximation B_k/M, and a noise of order $\gamma_\alpha^{-1}\sqrt{kM\log(M/\delta)/n}$. When $\operatorname{rank}\Theta^\star \leq 2k$, $B_k = 0$, and per-entry accuracy ε needs $n \gtrsim kM \log M/(\gamma_\alpha^2 \varepsilon^2)$ votes. Since each vote involves two agents, this is $\widetilde{O}(k/\varepsilon^2)$ votes per agent: the sampling cost grows with the complexity k, and not with the number of models M. The same corollary shows that the mean-square error per entry goes to zero as $\bar{N} \to \infty$.

Summary.

Term	Cause	Random	Bound	Subsection
(I) $\Theta_k^\dagger - \Theta^\star$	capacity k	no	Proposition 2	App. B.4.2
(II) $\Theta_k^{\dagger,\lambda} - \Theta_k^\dagger$	penalty λ	no	Proposition 3	App. B.4.3
(II) + (III) $\widehat{\Theta}_k - \Theta_k^\dagger$	penalty + finite votes	yes	Lemma 10 with Proposition 5	App. B.4.4
Total error	all three	yes	Theorem 1	App. B.4.4
Rate	balanced design	yes	Corollary B.1	App. B.5.1

B.2 Notation and observation model

Let $\mathbb{S}^K = \{X \in \mathbb{R}^{M\times M} : X^\top = -X\}$, with the trace inner product $\langle X, Y\rangle = \operatorname{Tr}(X^\top Y)$ and the induced Frobenius norm. We write $\|\cdot\|_*$ for the nuclear norm, $\|\cdot\|_{\mathrm{op}}$ for the operator norm, $\|X\|_{\max} = \max_{ij}|X_{ij}|$ for the entrywise sup-norm, and $\varsigma_1 \geq \varsigma_2 \geq \cdots$ for singular values. The rank of a real antisymmetric matrix is always even, so

$$\mathbb{D}_k = \{\Theta \in \mathbb{S}^K : \operatorname{rank}\Theta \leq 2k\}, \qquad \mathbb{D}_k(\alpha) = \mathbb{D}_k \cap \{\|\Theta\|_{\max} \leq \alpha\}$$

is exactly the disc-k class of Section 2. We fix

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \Omega_k = I_k \otimes J = \operatorname{blkdiag}(\underbrace{J, \ldots, J}_{k}),$$

and write the three properties used in our analysis:

$$\Omega_k^\top = -\Omega_k, \qquad \Omega_k^\top \Omega_k = I_{2k}, \qquad \|\Omega_k\|_{\mathrm{op}} = 1. \tag{7}$$

The first shows that $C\Omega_k C^\top$ is antisymmetric.

Throughout Θ denotes a *log-odds* matrix. The margin matrix defined in Section 2 is $A := \tanh(\Theta/2)$, entrywise.

Here (A1)–(A2) follow Section 3; (A3) denotes its independent Bernoulli observation model with finite votes and fixed q_{ij}, N_{ij} as described in Section 2, and (A4) its global-minimizer condition.

Lemma 1 (Transfer from logit to the margin scale)**.** *For all* $\Theta, \Theta' \in \mathbb{S}^K$,

$$\|A - A'\|_F \leq \tfrac{1}{2}\|\Theta - \Theta'\|_F \qquad \textit{and} \qquad \|A - A'\|_{\max} \leq \tfrac{1}{2}\|\Theta - \Theta'\|_{\max}. \tag{8}$$

Proof. The scalar map $x \mapsto \tanh(x/2)$ has derivative $\frac{1}{2}\operatorname{sech}^2(x/2)$, which lies in $(0, \frac{1}{2}]$ for all x. By the mean value theorem, for every entry $|A_{ij} - A'_{ij}| \leq \frac{1}{2}|\Theta_{ij} - \Theta'_{ij}|$. Taking the maximum over (i, j) gives the second claim; squaring and summing over (i, j) gives the first. □

Pair coordinates. A matrix in $\mathbb{S}^K$ supported on $\mathcal{E}$ is determined by its entries with $i < j$. All gradients are taken with respect to the *pair inner product*

$$\langle X, Y\rangle_{\mathcal{E}} := \sum_{(i,j)\in\mathcal{E}} X_{ij}Y_{ij}, \tag{9}$$

that is, $[\nabla f(\Theta)]_{ij} = \partial f/\partial\Theta_{ij}$ for $(i,j) \in \mathcal{E}$, extended antisymmetrically and by zero off $\mathcal{E}$. Two identities follow:

Lemma 2. *If $X, Y \in \mathbb{S}^K$ and either $\mathcal{E}$ is complete or one of X, Y vanishes off $\mathcal{E}$, then*

$$\langle X, Y\rangle_{\mathcal{E}} = \tfrac{1}{2}\langle X, Y\rangle, \qquad \|X\|_F^2 = 2\sum_{(i,j)\in\mathcal{E}} X_{ij}^2. \tag{10}$$

Proof. Antisymmetry gives $X_{ii} = 0$ and $X_{ji}Y_{ji} = (-X_{ij})(-Y_{ij}) = X_{ij}Y_{ij}$, so each unordered pair contributes twice to $\langle X, Y\rangle = \sum_{i,j} X_{ij}Y_{ij}$ and once to $\langle X, Y\rangle_{\mathcal{E}}$. Taking $Y = X$ gives the second identity. □

B.3 The estimator and population targets

Lemma 3 (Cross-entropy is the log-odds likelihood)**.** *With $\psi(x) = \log(1 + e^x)$, so that $\psi' = \sigma$ and $\psi'' = \sigma'$, for every $x \in \mathbb{R}$ and $y \in \{0, 1\}$,*

$$\mathrm{CE}\big(y, \sigma(x)\big) := -\big[y \log\sigma(x) + (1 - y)\log(1 - \sigma(x))\big] = -\log\mathbb{P}(y \mid x) = \psi(x) - yx.$$

Proof. Under the model $\mathbb{P}(y = 1 \mid x) = \sigma(x)$ we have $\mathbb{P}(y \mid x) = \sigma(x)^y(1 - \sigma(x))^{1-y}$, so

$$-\log\mathbb{P}(y \mid x) = -y\log\sigma(x) - (1 - y)\log(1 - \sigma(x)) = \mathrm{CE}\big(y, \sigma(x)\big),$$

which is the middle equality.

From $\sigma(x) = (1 + e^{-x})^{-1}$,

$$\log\sigma(x) = -\log(1 + e^{-x}), \qquad \log(1 - \sigma(x)) = \log\frac{e^{-x}}{1 + e^{-x}} = -x - \log(1 + e^{-x}).$$

Inserting Step 2 into Step 1,

$$-\log\mathbb{P}(y \mid x) = -y\big[-\log(1 + e^{-x})\big] - (1 - y)\big[-x - \log(1 + e^{-x})\big] = \log(1 + e^{-x}) + (1 - y)x.$$

$\log(1 + e^{-x}) + x = \log\big(e^x(1 + e^{-x})\big) = \log(1 + e^x) = \psi(x)$, so Step 3 equals $\psi(x) - yx$. □

Aggregation and sufficiency. Fix a pair $(i,j) \in \mathcal{E}$ and apply Lemma 3 at $x = \Theta_{ij}$ to each of its N_{ij} outcomes $y_{ij}^{(1)}, \ldots, y_{ij}^{(N_{ij})}$:

$$\sum_{t=1}^{N_{ij}} \mathrm{CE}\big(y_{ij}^{(t)}, \sigma(\Theta_{ij})\big) = \sum_{t=1}^{N_{ij}} \big[\psi(\Theta_{ij}) - y_{ij}^{(t)}\Theta_{ij}\big] = N_{ij}\psi(\Theta_{ij}) - W_{ij}\Theta_{ij},$$

since $W_{ij} = \sum_t y_{ij}^{(t)}$. Raw outcomes therefore never reappear: $\{W_{ij}\}$ is a sufficient statistic. Multiplying by the per-comparison weight q_{ij}/N_{ij} and summing over pairs gives the **empirical loss equation**:

$$\widehat{\mathcal{L}}(\Theta) = \sum_{(i,j)\in\mathcal{E}} q_{ij}\big[\psi(\Theta_{ij}) - \widehat{P}_{ij}\Theta_{ij}\big],$$

which is the loss of equation 3. Taking expectations and using $\mathbb{E}[\widehat{P}_{ij}] = P_{ij}^\star$, gives the **population loss equation**:

$$\overline{\mathcal{L}}(\Theta) = \sum_{(i,j)\in\mathcal{E}} q_{ij}\big[\psi(\Theta_{ij}) - P_{ij}^\star\Theta_{ij}\big], \qquad P_{ij}^\star = \sigma(\Theta_{ij}^\star) = \tfrac{1}{2}\big(1 + A_{ij}^\star\big).$$

Proposition 1 (Weight decay (L2 regularization) on C = nuclear norm on log-odds Θ)**.** *For every* $\Theta \in \mathbb{D}_k$*,*

$$\min\{\|C\|_F^2 : C\Omega_k C^\top = \Theta\} = \|\Theta\|_*. \tag{11}$$

Hence for any $\mathcal{S} \subseteq \mathbb{D}_k$*, any* $\lambda \geq 0$ *and any* f*,*

$$\min_{C:\, C\Omega_k C^\top \in \mathcal{S}} \Big\{ f(C\Omega_k C^\top) + \tfrac{\lambda}{2}\|C\|_F^2 \Big\} = \min_{\Theta \in \mathcal{S}} \Big\{ f(\Theta) + \tfrac{\lambda}{2}\|\Theta\|_* \Big\},$$

and $\widehat{\Theta} = \widehat{C}\Omega_k\widehat{C}^\top$ *solves the right-hand program whenever* $\widehat{C}$ *solves the left-hand one.*

Proof. Step 1. Write $\Theta = Q\,\mathrm{blkdiag}(\varsigma_l J)Q^\top$ as in Lemma 11 and set

$$C_\Theta := Q\ \mathrm{blkdiag}\big(\sqrt{\varsigma_1}I_2, \ldots, \sqrt{\varsigma_k}I_2\big) \in \mathbb{R}^{M\times 2k}.$$

Because $\sqrt{\varsigma_l}I_2$ is a scalar multiple of the identity on each 2×2 block, it commutes with J, and

$$C_\Theta \Omega_k C_\Theta^\top = Q\ \mathrm{blkdiag}(\sqrt{\varsigma_l}I_2)\ \mathrm{blkdiag}(J)\ \mathrm{blkdiag}(\sqrt{\varsigma_l}I_2)\, Q^\top = Q\ \mathrm{blkdiag}(\varsigma_l J)\, Q^\top = \Theta.$$

So C_Θ is feasible and this shows $C \mapsto C\Omega_k C^\top$ is *onto* $\mathbb{D}_k$, which is required.

Step 2. Using $Q^\top Q = I_{2k}$,

$$\|C_\Theta\|_F^2 = \mathrm{Tr}(C_\Theta^\top C_\Theta) = \mathrm{Tr}\big(\mathrm{blkdiag}(\sqrt{\varsigma_l}I_2)\ \underbrace{Q^\top Q}_{=I_{2k}}\ \mathrm{blkdiag}(\sqrt{\varsigma_l}I_2)\big) = \mathrm{Tr}\big(\mathrm{blkdiag}(\varsigma_l I_2)\big) = 2\sum_l \varsigma_l,$$

which equals $\|\Theta\|_*$ by equation 33. Hence the left side of equation 11 is at most $\|\Theta\|_*$.

Step 3 . Let C be any factor with $\Theta = C\Omega_k C^\top$ and let B satisfy $\|B\|_{\mathrm{op}} \leq 1$. Using $\Theta^\top = -\Theta$, then cyclicity of the trace, then $\Omega_k^\top = -\Omega_k$ from equation 7,

$$\begin{aligned}\langle \Theta, B\rangle &= \mathrm{Tr}(\Theta^\top B) = -\,\mathrm{Tr}(\Theta B) = -\,\mathrm{Tr}\big(C\Omega_k C^\top B\big)\\ &= -\,\mathrm{Tr}\big(\Omega_k\, C^\top B C\big) = -\langle \Omega_k^\top, C^\top B C\rangle = \langle \Omega_k, C^\top B C\rangle.\end{aligned}$$

Now apply Lemma 12(i) with $\|\Omega_k\|_{\mathrm{op}} = 1$, then Lemma 12(ii) with $X = C$ and $Y = BC$, then Lemma 12(iii):

$$\langle \Theta, B\rangle \ \leq\ \|\Omega_k\|_{\mathrm{op}}\|C^\top(BC)\|_* \ \leq\ \|C\|_F\|BC\|_F \ \leq\ \|B\|_{\mathrm{op}}\|C\|_F^2 \ \leq\ \|C\|_F^2.$$

Taking the supremum over $\|B\|_{\mathrm{op}} \leq 1$ and using Lemma 12(i) once more gives $\|\Theta\|_* \leq \|C\|_F^2$. With Step 2 this proves equation 11.

Step 4. Split the minimization over C into an outer minimization over $\Theta = C\Omega_k C^\top \in \mathcal{S}$ and an inner one over $\{C : C\Omega_k C^\top = \Theta\}$. The term f is constant on each C, and by equation 11 the inner minimization of $\frac{\lambda}{2}\|C\|_F^2$ returns $\frac{\lambda}{2}\|\Theta\|_*$. The statement about minimizers follows by tracking equality: if $\widehat{C}$ attains the left side then $\Theta = \widehat{C}\Omega_k\widehat{C}^\top$ attains the right, and conversely C_Θ attains the left whenever Θ attains the right. □

Remark 1 (On C). The C-program is non-convex. Note that the capabilities are not identifiable: $C\Omega_k C^\top$ is unchanged by $C \mapsto CS$ for every S in the set $\{S : S\Omega_k S^\top = \Omega_k\}$, of dimension $k(2k+1)$, so $\dim \mathbb{D}_k = 2Mk - k(2k+1)$. Every statement below is about Θ and A, never about C.

B.4 Error decomposition and finite-sample analysis

B.4.1 Shared likelihood identities and curvature

Lemma 4 (Derivatives)**.** *For* $(i,j) \in \mathcal{E}$*,*

$$[\nabla\widehat{\mathcal{L}}(\Theta)]_{ij} = q_{ij}\big(\sigma(\Theta_{ij}) - \widehat{P}_{ij}\big), \qquad [\nabla\overline{\mathcal{L}}(\Theta)]_{ij} = q_{ij}\big(\sigma(\Theta_{ij}) - P^\star_{ij}\big),$$

and $\nabla^2\widehat{\mathcal{L}}(\Theta) = \mathrm{diag}\big(q_{ij}\sigma'(\Theta_{ij})\big)$ *is diagonal in the pair basis. The likelihood is strictly convex in the observed pair coordinates. The rank constraint makes the feasible set nonconvex.*

Proof. Each summand of $\widehat{\mathcal{L}}(\Theta) = \sum_{\mathcal{E}} q_{ij}[\psi(\Theta_{ij}) - \widehat{P}_{ij}\Theta_{ij}]$ depends on exactly one coordinate Θ_{ij}, so $\partial\widehat{\mathcal{L}}/\partial\Theta_{ij} = q_{ij}[\psi'(\Theta_{ij}) - \widehat{P}_{ij}] = q_{ij}[\sigma(\Theta_{ij}) - \widehat{P}_{ij}]$ and all mixed second partials vanish, while $\partial^2\widehat{\mathcal{L}}/\partial\Theta_{ij}^2 = q_{ij}\psi''(\Theta_{ij}) = q_{ij}\sigma'(\Theta_{ij}) > 0$. The population version is identical with $\widehat{P}$ replaced by $P^\star$.

The antisymmetric extension is consistent with the model: evaluating the formula at (j, i) and using $q_{ji} = q_{ij}$, $\sigma(\Theta_{ji}) = \sigma(-\Theta_{ij}) = 1-\sigma(\Theta_{ij})$ and $\widehat{P}_{ji} = 1-\widehat{P}_{ij}$ gives $q_{ij}[(1-\sigma(\Theta_{ij})) - (1-\widehat{P}_{ij})] = -[\nabla\widehat{\mathcal{L}}(\Theta)]_{ij}$, so no symmetrization step is needed. □

Lemma 5. *Let $\mathcal{L}_z(\Theta) = \sum_{\mathcal{E}} q_{ij}[\psi(\Theta_{ij}) - z_{ij}\Theta_{ij}]$ for a fixed $z \in \mathbb{R}^{\mathcal{E}}$ (so $z = \widehat{P}$ gives $\widehat{\mathcal{L}}$ and $z = P^\star$ gives $\overline{\mathcal{L}}$). For any $\Theta^\circ, \Theta \in \mathbb{S}^K$, writing $\Upsilon = \Theta - \Theta^\circ$ and $S = \nabla\mathcal{L}_z(\Theta^\circ)$,*

$$\mathcal{L}_z(\Theta) - \mathcal{L}_z(\Theta^\circ) = \underbrace{\sum_{(i,j)\in\mathcal{E}} q_{ij} D_\psi\big(\Theta_{ij}, \Theta^\circ_{ij}\big)}_{\text{deterministic given } \Theta;\ \geq 0} + \underbrace{\langle S, \Upsilon\rangle_{\mathcal{E}}}_{\text{exactly linear in } \Upsilon}, \tag{12}$$

where $D_\psi(x, y) = \psi(x) - \psi(y) - \psi'(y)(x - y)$ is the Bregman divergence of ψ.

Proof. Step 1 . Directly from the definition,

$$\mathcal{L}_z(\Theta) - \mathcal{L}_z(\Theta^\circ) = \sum_{(i,j)\in\mathcal{E}} q_{ij}\Big\{\big[\psi(\Theta_{ij}) - \psi(\Theta^\circ_{ij})\big] - z_{ij}\Upsilon_{ij}\Big\}.$$

Step 2. Add an subtract $\pm\, q_{ij}\psi'(\Theta^\circ_{ij})\Upsilon_{ij}$ inside the brace and group:

$$= \sum_{(i,j)\in\mathcal{E}} q_{ij} \underbrace{\big[\psi(\Theta_{ij}) - \psi(\Theta^\circ_{ij}) - \psi'(\Theta^\circ_{ij})\Upsilon_{ij}\big]}_{=D_\psi(\Theta_{ij},\Theta^\circ_{ij})} + \sum_{(i,j)\in\mathcal{E}} q_{ij} \underbrace{\big[\psi'(\Theta^\circ_{ij}) - z_{ij}\big]}_{=S_{ij}/q_{ij}\cdot q_{ij}} \Upsilon_{ij}.$$

Step 3 By Lemma 4, $S_{ij} = q_{ij}(\sigma(\Theta^\circ_{ij}) - z_{ij}) = q_{ij}(\psi'(\Theta^\circ_{ij}) - z_{ij})$, so the second sum is $\sum_{\mathcal{E}} S_{ij}\Upsilon_{ij} = \langle S, \Upsilon\rangle_{\mathcal{E}}$ by equation 9. Nonnegativity of the first sum is convexity of ψ. □

Lemma 6 (Lower and upper bounds for D_ψ)**.** *For $\alpha \geq 0$, $\gamma_\alpha = \sigma'(\alpha) \in [\frac{1}{4}e^{-\alpha}, \frac{1}{4}]$, and if $|x|, |y| \leq \alpha$ then*

$$\tfrac{\gamma_\alpha}{2}(x - y)^2 \;\leq\; D_\psi(x, y) \;\leq\; \tfrac{1}{8}(x - y)^2. \tag{13}$$

Proof. Step 1 . $\sigma'(z) = \sigma(z)(1-\sigma(z)) = e^{-z}/(1+e^{-z})^2$ is strictly positive, even in z, maximized at $z = 0$ with $\sigma'(0) = \frac{1}{4}$, and strictly decreasing on $[0, \infty)$. Hence $|z| \leq \alpha$ implies $\gamma_\alpha = \sigma'(\alpha) \leq \sigma'(z) \leq \frac{1}{4}$.

Step 2 (the lower bound on γ_α). With $\gamma_\alpha = e^\alpha/(1 + e^\alpha)^2$,

$$\gamma_\alpha \geq \tfrac{1}{4}e^{-\alpha} \iff 4e^{2\alpha} \geq (1 + e^\alpha)^2 \iff 2e^\alpha \geq 1 + e^\alpha \iff e^\alpha \geq 1,$$

which holds for every $\alpha \geq 0$.

Step 3 (Taylor with Lagrange remainder). Since ψ is twice differentiable, there is ξ strictly between x and y with

$$D_\psi(x, y) = \psi(x) - \psi(y) - \psi'(y)(x - y) = \tfrac{1}{2}\psi''(\xi)(x - y)^2 = \tfrac{1}{2}\sigma'(\xi)(x - y)^2.$$

As ξ lies between x and y we have $|\xi| \leq \max(|x|, |y|) \leq \alpha$, so Step 1 gives $\frac{1}{2}\sigma'(\xi) \in [\frac{\gamma_\alpha}{2}, \frac{1}{8}]$, which is equation 13. □

Lemma 7 (Noise/bias split of the score)**.** *For every $\Theta^\circ \in \mathbb{S}^K$,*

$$\nabla\widehat{\mathcal{L}}(\Theta^\circ) = Z + \nabla\overline{\mathcal{L}}(\Theta^\circ), \qquad \textit{where} \qquad Z_{ij} = q_{ij}\big(P^\star_{ij} - \widehat{P}_{ij}\big), \tag{14}$$

and Z does not *depend on Θ°: the stochastic part of the score is the same matrix wherever the gradient is evaluated. Moreover $\nabla\overline{\mathcal{L}}(\Theta^\star) = 0$ and, for complete $\mathcal{E}$,*

$$\|\nabla\overline{\mathcal{L}}(\Theta)\|_{\mathrm{op}} \;\leq\; \|\nabla\overline{\mathcal{L}}(\Theta)\|_F \;\leq\; \tfrac{q_{\max}}{4}\|\Theta - \Theta^\star\|_F. \tag{15}$$

Proof. Step 1. Subtract the two gradients (empirical and population) of Lemma 4. The terms $q_{ij}\sigma(\Theta^{\circ}_{ij})$ cancel and what remains is $q_{ij}(P^{\star}_{ij} - \widehat{P}_{ij}) = Z_{ij}$, in which Θ° does not appear.

Step 2 (the population scores is 0 at the truth). By definition $P^{\star}_{ij} = \sigma(\Theta^{\star}_{ij})$, so $[\nabla\overline{\mathcal{L}}(\Theta^{\star})]_{ij} = q_{ij}(\sigma(\Theta^{\star}_{ij}) - P^{\star}_{ij}) = 0$.

Step 3. Since $\sigma' \le \frac{1}{4}$ everywhere, σ is $\frac{1}{4}$-Lipschitz, so by Step 2

$$\big|[\nabla\overline{\mathcal{L}}(\Theta)]_{ij}\big| = q_{ij}\big|\sigma(\Theta_{ij}) - \sigma(\Theta^{\star}_{ij})\big| \le \tfrac{q_{ij}}{4}\big|\Theta_{ij} - \Theta^{\star}_{ij}\big|.$$

Squaring, summing over $(i,j) \in \mathcal{E}$ and multiplying by 2 (Lemma 2),

$$\|\nabla\overline{\mathcal{L}}(\Theta)\|_F^2 = 2\sum_{(i,j)\in\mathcal{E}}[\nabla\overline{\mathcal{L}}(\Theta)]_{ij}^2 \le \frac{q_{\max}^2}{16}\cdot 2\sum_{(i,j)\in\mathcal{E}}(\Theta_{ij} - \Theta^{\star}_{ij})^2 = \frac{q_{\max}^2}{16}\|\Theta - \Theta^{\star}\|_F^2,$$

and applying $\|\cdot\|_{\mathrm{op}} \le \|\cdot\|_F$. □

Lemma 8. *For every $\Theta \in \mathbb{S}^K$, $\overline{\mathcal{L}}(\Theta) - \overline{\mathcal{L}}(\Theta^{\star}) = \sum_{\mathcal{E}} q_{ij} D_{\psi}(\Theta_{ij}, \Theta^{\star}_{ij}) \ge 0$, with equality if and only if $\Theta = \Theta^{\star}$ on $\mathcal{E}$. Hence $\Theta^{\star}$ minimizes $\overline{\mathcal{L}}$, uniquely on the observed support — so uniqueness of the population target requires the pair set to be complete.*

Proof. Apply Lemma 5 with $z = P^{\star}$ and $\Theta^{\circ} = \Theta^{\star}$. The linear term vanishes because $S = \nabla\overline{\mathcal{L}}(\Theta^{\star}) = 0$ by Lemma 7, leaving the Bregman sum. Since ψ is strictly convex, $D_{\psi}(x,y) > 0$ whenever $x \ne y$, and $q_{ij} > 0$. This means the sum is zero exactly when $\Theta_{ij} = \Theta^{\star}_{ij}$ for all $(i,j) \in \mathcal{E}$. Because entries outside $\mathcal{E}$ are unconstrained, we get the uniqueness problem. □

Remark 2 (Why should disc structure use log-odds). Suppose one instead imposes $\operatorname{rank} A \le 2k$ on the *margin* matrix and penalizes $\|A\|_*$. The per-comparison negative log-likelihood in that parameterization is, up to constants, $-\log(1 - A_{ij}) - 2y\,\operatorname{artanh}(A_{ij})$, still *affine in the outcome* but with coefficient $2\operatorname{artanh}(A_{ij}) = \Theta_{ij}$. Repeating the computation above therefore leaves the noise term in the form $\langle Z, \Theta - \Theta^{\circ}\rangle_{\mathcal{E}}$ — paired against the *log-odds* error, not against the margin error $\Upsilon := A - A^{\circ}$ that the penalty and the rank constraint now control. Note this is a remark about the *analysis* and whether A or Θ is the more natural carrier of cyclic structure is a modelling question, and the two classes differ.

B.4.2 Approximation error (I)

Proposition 2 (Approximation term)**.** *Let Θ^{tr}_k be the rank-$2k$ Youla truncation of $\Theta^{\star}$ and suppose $\Theta^{\mathrm{tr}}_k \in \mathbb{D}_k(\alpha)$. Then under (A1),*

$$\sum_{\mathcal{E}} q_{ij}\big(\Theta^{\dagger}_{k,ij} - \Theta^{\star}_{ij}\big)^2 \;\le\; \frac{1}{4\gamma_{\alpha}}\sum_{\mathcal{E}} q_{ij}\big(\Theta^{\mathrm{tr}}_{k,ij} - \Theta^{\star}_{ij}\big)^2, \tag{16}$$

and if $\mathcal{E}$ is complete then

$$B_k \;\le\; \frac{1}{2\sqrt{\gamma_{\alpha}}}\sqrt{\frac{q_{\max}}{q_{\min}}}\,\varsigma_{>k} \;\le\; e^{\alpha/2}\sqrt{\frac{q_{\max}}{q_{\min}}}\,\varsigma_{>k}, \qquad \varsigma_{>k} := \|\Theta^{\mathrm{tr}}_k - \Theta^{\star}\|_F = \Big(2\sum_{l>k}\varsigma_l(\Theta^{\star})^2\Big)^{1/2}.$$

Proof. Step 1 (lower bound the excess risk at $\Theta^{\dagger}_k$). Both $\Theta^{\dagger}_k$ and $\Theta^{\star}$ lie in $\mathbb{D}_k(\alpha)$, so Lemma 8 and the *lower* half of equation 13 give

$$\overline{\mathcal{L}}(\Theta^{\dagger}_k) - \overline{\mathcal{L}}(\Theta^{\star}) = \sum_{\mathcal{E}} q_{ij} D_{\psi}\big(\Theta^{\dagger}_{k,ij}, \Theta^{\star}_{ij}\big) \;\ge\; \frac{\gamma_{\alpha}}{2}\sum_{\mathcal{E}} q_{ij}\big(\Theta^{\dagger}_{k,ij} - \Theta^{\star}_{ij}\big)^2. \tag{17}$$

Step 2. By hypothesis $\Theta^{\mathrm{tr}}_k \in \mathbb{D}_k(\alpha)$ is feasible, and $\Theta^{\dagger}_k$ minimizes $\overline{\mathcal{L}}$ over that set, so

$$\overline{\mathcal{L}}(\Theta^{\dagger}_k) - \overline{\mathcal{L}}(\Theta^{\star}) \;\le\; \overline{\mathcal{L}}(\Theta^{\mathrm{tr}}_k) - \overline{\mathcal{L}}(\Theta^{\star}). \tag{18}$$

Step 3 (upper bound the excess risk at the truncation). By Lemma 8 and the *upper* half of equation 13,

$$\overline{\mathcal{L}}(\Theta_k^{\mathrm{tr}}) - \overline{\mathcal{L}}(\Theta^\star) \;\le\; \frac{1}{8}\sum_{\mathcal{E}} q_{ij}\big(\Theta_{k,ij}^{\mathrm{tr}} - \Theta_{ij}^\star\big)^2. \tag{19}$$

Step 4. Combining equation 17, equation 18 and equation 19,

$$\frac{\gamma_\alpha}{2}\sum_{\mathcal{E}} q_{ij}\big(\Theta_{k,ij}^{\dagger} - \Theta_{ij}^\star\big)^2 \;\le\; \frac{1}{8}\sum_{\mathcal{E}} q_{ij}\big(\Theta_{k,ij}^{\mathrm{tr}} - \Theta_{ij}^\star\big)^2,$$

and multiplying by $2/\gamma_\alpha$ gives equation 16.

Step 5. Bound the left side of equation 16 below by $q_{\min}$ times the unweighted sum and the right side above by $q_{\max}$ times the unweighted sum:

$$\sum_{i<j}\big(\Theta_{k,ij}^{\dagger} - \Theta_{ij}^\star\big)^2 \;\le\; \frac{q_{\max}}{4\gamma_\alpha q_{\min}}\sum_{i<j}\big(\Theta_{k,ij}^{\mathrm{tr}} - \Theta_{ij}^\star\big)^2.$$

When $\mathcal{E}$ is complete, both sums run over *all* pairs $i<j$, so multiplying by 2 turns each into a Frobenius norm by Lemma 2: $B_k^2 \le \frac{q_{\max}}{4\gamma_\alpha q_{\min}}\varsigma_{>k}^2$. Taking square roots gives the first bound, and $\gamma_\alpha \ge \frac{1}{4}e^{-\alpha}$ (Lemma 6) gives $(2\sqrt{\gamma_\alpha})^{-1} \le e^{\alpha/2}$, which is the second. □

B.4.3 Regularization bias (II) and a shared deterministic bound

Fix a reference matrix $\Theta^\circ \in \mathbb{D}_k$ with Youla basis Q (Lemma 11) and let $P_Q = QQ^\top$ be the orthogonal projector onto $\mathrm{range}(Q)$. For $\Upsilon \in \mathbb{S}^K$ split

$$\Upsilon' = P_Q\Upsilon + \Upsilon P_Q - P_Q\Upsilon P_Q, \qquad \Upsilon'' = (I-P_Q)\Upsilon(I-P_Q). \tag{20}$$

Lemma 9 (Split and decomposability)**.** *For every $\Upsilon \in \mathbb{S}^K$:*

(i) $\Upsilon = \Upsilon' + \Upsilon''$ *with* $\Upsilon', \Upsilon'' \in \mathbb{S}^K$ *and* $\langle \Upsilon', \Upsilon''\rangle = 0$*; hence* $\|\Upsilon'\|_F \le \|\Upsilon\|_F$.

(ii) $\mathrm{rank}\,\Upsilon' \le 4k$.

(iii) $\|\Theta^\circ + \Upsilon''\|_* = \|\Theta^\circ\|_* + \|\Upsilon''\|_*$.

(iv) $\|\Theta^\circ + \Upsilon\|_* \;\ge\; \|\Theta^\circ\|_* + \|\Upsilon''\|_* - \|\Upsilon'\|_*$.

Proof. (i) Adding the Υ' and Υ'' in equation 20 gives $\Upsilon' + \Upsilon'' = \Upsilon$. Since P_Q is symmetric and $\Upsilon^\top = -\Upsilon$,

$$(\Upsilon')^\top = \Upsilon^\top P_Q + P_Q\Upsilon^\top - P_Q\Upsilon^\top P_Q = -\Upsilon',$$

so $\Upsilon' \in \mathbb{S}^K$, and likewise for Υ''. The map $\Upsilon \mapsto \Upsilon'$ is the orthogonal projector onto $T = \{Y : (I-P_Q)Y(I-P_Q) = 0\}$, so the split is orthogonal and each piece has Frobenius norm at most that of Υ.

(ii) Rewrite $\Upsilon' = P_Q\Upsilon + (I-P_Q)\Upsilon P_Q$. The first term has column space inside $\mathrm{range}(P_Q)$, of dimension $2k$. The second has row space inside $\mathrm{range}(P_Q)$ of dimension $2k$. Hence $\mathrm{rank}\,\Upsilon' \le 2k + 2k = 4k$.

(iii) $\mathrm{col}(\Theta^\circ) \subseteq \mathrm{range}(P_Q)$ while $\mathrm{col}(\Upsilon'') \subseteq \mathrm{range}(I-P_Q)$, and the same holds for the row spaces. For two matrices whose row spaces and column spaces are mutually orthogonal the singular values of the sum are the union of the two singular value lists, so the nuclear norms add.

(iv) By the triangle inequality and then (iii),

$$\|\Theta^\circ + \Upsilon\|_* \;\ge\; \|\Theta^\circ + \Upsilon''\|_* - \|\Upsilon'\|_* \;=\; \|\Theta^\circ\|_* + \|\Upsilon''\|_* - \|\Upsilon'\|_*. \qquad \square$$

Definition 1 (Error cone assumption (A2) in deatil)**.** Let $\mathbb{C}(\Theta^\circ) := \{\Upsilon \in \mathbb{S}^K : \|\Upsilon''\|_* \le 3\|\Upsilon'\|_*\}$, the split taken relative to Θ°. Assumption (A2) states that there is $\nu > 0$ with

$$\sum_{(i,j)\in\mathcal{E}} q_{ij}\Upsilon_{ij}^2 \;\ge\; \tfrac{\nu}{2}\|\Upsilon\|_F^2 \qquad \text{for every } \Upsilon \in \mathbb{C}(\Theta^\circ) \text{ with } \|\Upsilon\|_{\max} \le 2\alpha. \tag{21}$$

For a complete pair set with $q_{ij} \equiv \bar{q}$ the left side equals $\bar{q}\sum_{i<j}\Upsilon_{ij}^2 = \frac{\bar{q}}{2}\|\Upsilon\|_F^2$ by Lemma 2, so equation 21 holds with $\nu = \bar{q}$ *identically*; for general weights on a complete set it holds with $\nu = q_{\min}$. The assumption is useful for sparse or unbalanced matchmaking schedules.

Lemma 10 (Master lemma)**.** *Let $\mathcal{L}_z$ be as in Lemma 5, let $\Theta^\circ \in \mathbb{D}_k(\alpha)$, and let $\Theta^\bullet$ minimize $\mathcal{L}_z(\Theta) + \frac{\lambda}{2}\|\Theta\|_*$ over $\mathbb{D}_k(\alpha)$. Put $\Upsilon = \Theta^\bullet - \Theta^\circ$ and $S = \nabla\mathcal{L}_z(\Theta^\circ)$. If $\lambda \geq 2\|S\|_{\mathrm{op}}$ then $\Upsilon \in \mathbb{C}(\Theta^\circ)$, and under (A2),*

$$\|\Theta^\bullet - \Theta^\circ\|_F \;\leq\; \frac{6\sqrt{k}\,\lambda}{\gamma_\alpha\,\nu}. \tag{22}$$

Proof. Step 1. Θ° is feasible and $\Theta^\bullet$ is a global minimizer over the same set, so

$$\mathcal{L}_z(\Theta^\bullet) + \tfrac{\lambda}{2}\|\Theta^\bullet\|_* \;\leq\; \mathcal{L}_z(\Theta^\circ) + \tfrac{\lambda}{2}\|\Theta^\circ\|_*.$$

Rearranging and substituting the identity equation 12 for $\mathcal{L}_z(\Theta^\bullet) - \mathcal{L}_z(\Theta^\circ)$,

$$\mathrm{LHS} := \sum_{(i,j)\in\mathcal{E}} q_{ij} D_\psi\big(\Theta^\bullet_{ij}, \Theta^\circ_{ij}\big) \;\leq\; \underbrace{-\langle S, \Upsilon\rangle_{\mathcal{E}}}_{\text{noise}} + \underbrace{\tfrac{\lambda}{2}\big(\|\Theta^\circ\|_* - \|\Theta^\bullet\|_*\big)}_{\text{penalty}}. \tag{23}$$

Step 2. Both $\Theta^\bullet$ and Θ° lie in $\mathbb{D}_k(\alpha)$, so $|\Theta^\bullet_{ij}| \leq \alpha$ and $|\Theta^\circ_{ij}| \leq \alpha$ and the lower half of equation 13 applies to each summand:

$$\mathrm{LHS} \;\geq\; \frac{\gamma_\alpha}{2} \sum_{(i,j)\in\mathcal{E}} q_{ij}\Upsilon_{ij}^2 \;\geq\; 0. \tag{24}$$

Step 3. Since S vanishes off $\mathcal{E}$, Lemma 2 gives $\langle S, \Upsilon\rangle_{\mathcal{E}} = \frac{1}{2}\langle S, \Upsilon\rangle$. Hence, by Lemma 12(i), the hypothesis $\lambda \geq 2\|S\|_{\mathrm{op}}$, and the triangle inequality $\|\Upsilon\|_* \leq \|\Upsilon'\|_* + \|\Upsilon''\|_*$,

$$-\langle S, \Upsilon\rangle_{\mathcal{E}} = -\tfrac{1}{2}\langle S, \Upsilon\rangle \;\leq\; \tfrac{1}{2}\|S\|_{\mathrm{op}}\|\Upsilon\|_* \;\leq\; \tfrac{\lambda}{4}\|\Upsilon\|_* \;\leq\; \tfrac{\lambda}{4}\big(\|\Upsilon'\|_* + \|\Upsilon''\|_*\big). \tag{25}$$

Nuclearoperator norm duality gives the bound score $\|Z\|_{\mathrm{op}}$.

Step 4. Since $\operatorname{rank}\Theta^\circ \leq 2k$, Lemma 9(iv) applied to $\Theta^\bullet = \Theta^\circ + \Upsilon$ gives $\|\Theta^\bullet\|_* \geq \|\Theta^\circ\|_* + \|\Upsilon''\|_* - \|\Upsilon'\|_*$, i.e.

$$\tfrac{\lambda}{2}\big(\|\Theta^\circ\|_* - \|\Theta^\bullet\|_*\big) \;\leq\; \tfrac{\lambda}{2}\big(\|\Upsilon'\|_* - \|\Upsilon''\|_*\big). \tag{26}$$

Step 5. Insert equation 25 and equation 26 into equation 23 and collect the two nuclear norms:

$$\mathrm{LHS} \;\leq\; \Big(\tfrac{\lambda}{4} + \tfrac{\lambda}{2}\Big)\|\Upsilon'\|_* + \Big(\tfrac{\lambda}{4} - \tfrac{\lambda}{2}\Big)\|\Upsilon''\|_* \;=\; \tfrac{3\lambda}{4}\|\Upsilon'\|_* - \tfrac{\lambda}{4}\|\Upsilon''\|_*. \tag{27}$$

By equation 24 the left side is nonnegative, so $\frac{\lambda}{4}\|\Upsilon''\|_* \leq \frac{3\lambda}{4}\|\Upsilon'\|_*$, i.e. $\|\Upsilon''\|_* \leq 3\|\Upsilon'\|_*$ and $\Upsilon \in \mathbb{C}(\Theta^\circ)$. The preceding argument establishes the cone constraint before invoking (A2).

Step 6. Now (A2) applies to Υ. Use equation 24, then equation 21, then equation 27, then Lemma 12(iv) with $\operatorname{rank}\Upsilon' \leq 4k$ and $\|\Upsilon'\|_F \leq \|\Upsilon\|_F$ from Lemma 9:

$$\frac{\gamma_\alpha\nu}{4}\|\Upsilon\|_F^2 \overset{\text{equation 21}}{\leq} \frac{\gamma_\alpha}{2}\sum_{\mathcal{E}} q_{ij}\Upsilon_{ij}^2 \overset{\text{equation 24}}{\leq} \mathrm{LHS}$$

$$\overset{\text{equation 27}}{\leq} \frac{3\lambda}{4}\|\Upsilon'\|_* \;\leq\; \frac{3\lambda}{4}\sqrt{4k}\,\|\Upsilon'\|_F \;\leq\; \frac{3\lambda\sqrt{k}}{2}\|\Upsilon\|_F.$$

If $\Upsilon = 0$ the claim is trivial. Otherwise divide both ends by $\|\Upsilon\|_F$ and then by $\gamma_\alpha\nu/4$:

$$\|\Upsilon\|_F \;\leq\; \frac{4}{\gamma_\alpha\nu}\cdot\frac{3\lambda\sqrt{k}}{2} \;=\; \frac{6\sqrt{k}\,\lambda}{\gamma_\alpha\nu},$$

which is equation 22. □

Remark 3. The bound equation 22 is deterministic. All randomness is confined to the *event* $\lambda \geq 2\|S\|_{\mathrm{op}}$. The three parts: $\sqrt{k}$ from the rank–norm inequality on the tangent part of the error, $\gamma_\alpha^{-1} = O(e^\alpha)$ from the curvature floor and assumptions, and ν^{-1} from the design/matchmaking schedule. The bound holds without an asymptotic limit, under the hypotheses of Lemma 10.

Proposition 3 (The regularization term (II) alone)**.** *Let $B = \nabla\overline{\mathcal{L}}(\Theta_k^\dagger)$, so that $\|B\|_{\mathrm{op}} \leq \frac{q_{\max}}{4}B_k$ by equation 15, and $B = 0$ when the model is well specified. If $\lambda \geq 2\|B\|_{\mathrm{op}}$ then, under (A1)–(A3),*

$$\|\Theta_k^{\dagger,\lambda} - \Theta_k^\dagger\|_F \;\leq\; \frac{6\sqrt{k}\,\lambda}{\gamma_\alpha\nu}.$$

In the well-specified case the hypothesis is vacuous and the bound holds for every $\lambda \geq 0$.

Proof. Apply Lemma 10 with $\mathcal{L}_z = \overline{\mathcal{L}}$ (that is, $z = P^\star$), $\Theta^\circ = \Theta_k^\dagger$ and $\Theta^\bullet = \Theta_k^{\dagger,\lambda}$. These are, by definition, the unpenalized and the penalized minimizers of the *same* population risk over the *same* feasible set $\mathbb{D}_k(\alpha)$, so both hypotheses of the master lemma hold, and $S = \nabla\overline{\mathcal{L}}(\Theta_k^\dagger) = B$. Well specification gives $\Theta_k^\dagger = \Theta^\star$, hence $B = \nabla\overline{\mathcal{L}}(\Theta^\star) = 0$ by Lemma 7. □

B.4.4 Sampling error (III) and the joint finite-sample bound

Proposition 4 (Moments of the score)**.** *Under (A3) the variables $\{Z_{ij}\}_{(i,j)\in\mathcal{E}}$ are independent and mean-zero, with*

$$\mathrm{Var}(Z_{ij}) = \frac{q_{ij}^2\,\sigma'(\Theta_{ij}^\star)}{N_{ij}} \le \frac{q_{ij}^2}{4N_{ij}}, \qquad |Z_{ij}| \le q_{ij}, \tag{28}$$

and Z_{ij} is sub-Gaussian with proxy $q_{ij}/(2\sqrt{N_{ij}})$.

Proof. Step 1. Since $\widehat{P}_{ij} = N_{ij}^{-1}\sum_t y_{ij}^{(t)}$,

$$Z_{ij} = q_{ij}\big(P_{ij}^\star - \widehat{P}_{ij}\big) = -\frac{q_{ij}}{N_{ij}}\sum_{t=1}^{N_{ij}}\big(y_{ij}^{(t)} - P_{ij}^\star\big), \tag{29}$$

a scaled sum of N_{ij} i.i.d. centered Bernoulli variables.

Step 2. Under (A3) the counts are fixed in advance and outcomes are independent across pairs, so distinct Z_{ij} involve disjoint sets of comparisons and are independent; $\mathbb{E}[y_{ij}^{(t)}] = P_{ij}^\star$ gives $\mathbb{E}[Z_{ij}] = 0$.

Step 3. From equation 29,

$$\mathrm{Var}(Z_{ij}) = \frac{q_{ij}^2}{N_{ij}^2}\cdot N_{ij}\,P_{ij}^\star(1-P_{ij}^\star) = \frac{q_{ij}^2 P_{ij}^\star(1-P_{ij}^\star)}{N_{ij}} = \frac{q_{ij}^2\sigma'(\Theta_{ij}^\star)}{N_{ij}},$$

using $P_{ij}^\star = \sigma(\Theta_{ij}^\star)$ and $\sigma' = \sigma(1-\sigma)$. The bound $\sigma' \le \frac{1}{4}$ gives the stated inequality.

Step 4. Since $\widehat{P}_{ij} \in [0,1]$ and $P_{ij}^\star \in (0,1)$, $|P_{ij}^\star - \widehat{P}_{ij}| \le \max\{P_{ij}^\star, 1-P_{ij}^\star\} \le 1$, so $|Z_{ij}| \le q_{ij}$.

Step 5. Each $y_{ij}^{(t)} - P_{ij}^\star$ lies in an interval of length 1, hence has proxy $\frac{1}{2}$ by Lemma 14; the sum of N_{ij} of them has proxy $\sqrt{N_{ij}}/2$, and multiplying by q_{ij}/N_{ij} gives $\frac{q_{ij}}{N_{ij}}\cdot\frac{\sqrt{N_{ij}}}{2} = \frac{q_{ij}}{2\sqrt{N_{ij}}}$. □

Proposition 4 is the consistency statement: because the projection weights q are fixed in advance, $\mathrm{Var}(Z_{ij}) \to 0$ as $N_{ij} \to \infty$.

Proposition 5 (Operator norm of the score)**.** *Under (A1), (A3), with probability at least $1-\delta$,*

$$\|Z\|_{\mathrm{op}} \;\le\; \zeta_\delta := \sqrt{2d_\star \log\tfrac{2M}{\delta}} + \tfrac{2}{3}\mathsf{q}_\star \log\tfrac{2M}{\delta}, \tag{30}$$

where $d_\star = \max_i \sum_j q_{ij}^2\sigma'(\Theta_{ij}^\star)/N_{ij}$ and $\mathsf{q}_\star = \max_{\mathcal{E}} q_{ij}/N_{ij}$.

Proof. Step 1. Write Z as a sum over individual comparisons. Let E_{ab} denote the matrix units and $S_{ij} := E_{ij} - E_{ji}$. By equation 29,

$$Z = \sum_{(i,j)\in\mathcal{E}}\sum_{t=1}^{N_{ij}} Y_{ij,t}, \qquad Y_{ij,t} := -\frac{q_{ij}}{N_{ij}}\big(y_{ij}^{(t)} - P_{ij}^\star\big)S_{ij}.$$

The summands are indexed by *single comparisons* and are independent by (A3).

Step 2. Each $Y_{ij,t}$ is real antisymmetric, hence not Hermitian, so we cannot apply Lemma 13 directly. Put $X_{ij,t} := \mathrm{i}\,Y_{ij,t}$. Then $X_{ij,t}^* = -\mathrm{i}\,Y_{ij,t}^\top = \mathrm{i}\,Y_{ij,t} = X_{ij,t}$, so $X_{ij,t}$ is Hermitian, and $\|\sum X_{ij,t}\|_{\mathrm{op}} = \|\mathrm{i}Z\|_{\mathrm{op}} = \|Z\|_{\mathrm{op}}$.

Step 3 (uniform bound R). S_{ij} has singular values $1, 1, 0, \ldots, 0$, so $\|S_{ij}\|_{\mathrm{op}} = 1$ and

$$\|X_{ij,t}\|_{\mathrm{op}} = \frac{q_{ij}}{N_{ij}} \left| y_{ij}^{(t)} - P_{ij}^{\star} \right| \;\le\; \frac{q_{ij}}{N_{ij}} \;\le\; \mathsf{q}_{\star} =: R.$$

Step 4 (variance parameter v). Using $E_{ab}E_{cd} = \delta_{bc}E_{ad}$ and $i \neq j$,

$$S_{ij}^2 = (E_{ij} - E_{ji})^2 = \underbrace{E_{ij}E_{ij}}_{=0} - \underbrace{E_{ij}E_{ji}}_{=E_{ii}} - \underbrace{E_{ji}E_{ij}}_{=E_{jj}} + \underbrace{E_{ji}E_{ji}}_{=0} = -(E_{ii} + E_{jj}).$$

Therefore $X_{ij,t}^2 = -Y_{ij,t}^2 = \frac{q_{ij}^2}{N_{ij}^2}(y_{ij}^{(t)} - P_{ij}^{\star})^2(E_{ii} + E_{jj})$, and taking expectations and summing over $t = 1, \ldots, N_{ij}$,

$$\sum_t \mathbb{E}X_{ij,t}^2 = N_{ij} \cdot \frac{q_{ij}^2 \, P_{ij}^{\star}(1 - P_{ij}^{\star})}{N_{ij}^2}(E_{ii} + E_{jj}) = \frac{q_{ij}^2 \, \sigma'(\Theta_{ij}^{\star})}{N_{ij}} \, (E_{ii} + E_{jj}).$$

Summing over pairs gives a nonnegative *diagonal* matrix whose i-th diagonal entry is $\sum_j q_{ij}^2 \sigma'(\Theta_{ij}^{\star})/N_{ij}$, the sum running over j with $(i, j) \in \mathcal{E}$. The operator norm of a nonnegative diagonal matrix is its largest entry, so

$$v = \Big\| \sum_{(i,j)\in\mathcal{E}} \sum_t \mathbb{E}X_{ij,t}^2 \Big\|_{\mathrm{op}} = \max_i \sum_j \frac{q_{ij}^2 \sigma'(\Theta_{ij}^{\star})}{N_{ij}} = d_{\star}.$$

Step 5 (apply Bernstein). Insert $R = \mathsf{q}_{\star}$ and $v = d_{\star}$ into Lemma 13. □

Remark 4 (What $d_{\star}$ says about scheduling?). The bound depends on the maximum row variance ($d_{\star}$) and the largest per-vote weight ($q_{\star}$). With reference weights fixed, each pairs contribution decreases as its vote count increases. Pairs assigned substantial weight but few comparisons can therefore dominate the bound. Under a fixed total budget, allocating more votes to some pairs reduces the votes available for others and degrades their estimates.

Proof of Theorem 1. Step 1. Apply Lemma 10 with $\mathcal{L}_z = \widehat{\mathcal{L}}$, $\Theta^{\circ} = \Theta_k^{\dagger}$ and $\Theta^{\bullet} = \widehat{\Theta}_k$. The hypotheses hold: $\Theta_k^{\dagger} \in \mathbb{D}_k(\alpha)$ has rank at most $2k$ by definition, and $\widehat{\Theta}_k$ is a global minimizer of $\widehat{\mathcal{L}} + \frac{\lambda}{2}\|\cdot\|_*$ over $\mathbb{D}_k(\alpha)$ by (A4).

Step 2. By Lemma 7, $S = \nabla\widehat{\mathcal{L}}(\Theta_k^{\dagger}) = Z + \nabla\overline{\mathcal{L}}(\Theta_k^{\dagger})$, so by the triangle inequality and equation 15,

$$\|S\|_{\mathrm{op}} \;\le\; \|Z\|_{\mathrm{op}} + \|\nabla\overline{\mathcal{L}}(\Theta_k^{\dagger})\|_{\mathrm{op}} \;\le\; \|Z\|_{\mathrm{op}} + \tfrac{q_{\max}}{4} B_k.$$

Step 3. On the event of Proposition 5, which has probability at least $1 - \delta$, we have $\|Z\|_{\mathrm{op}} \le \zeta_{\delta}$ and therefore $2\|S\|_{\mathrm{op}} \le 2\zeta_{\delta} + \frac{q_{\max}}{2} B_k = \lambda_{\min}$. Any $\lambda \ge \lambda_{\min}$ thus satisfies the hypothesis $\lambda \ge 2\|S\|_{\mathrm{op}}$ of Lemma 10.

Step 4. equation 22 gives $\|\widehat{\Theta}_k - \Theta_k^{\dagger}\|_F \le 6\sqrt{k}\lambda/(\gamma_{\alpha}\nu)$, so

$$\|\widehat{\Theta}_k - \Theta^{\star}\|_F \;\le\; \|\Theta_k^{\dagger} - \Theta^{\star}\|_F + \|\widehat{\Theta}_k - \Theta_k^{\dagger}\|_F \;\le\; B_k + \frac{6\sqrt{k}}{\gamma_{\alpha}\nu}\lambda,$$

which is the first bound in Theorem 1.

Step 5. Substituting $\lambda = \lambda_{\min} = 2\zeta_{\delta} + \frac{q_{\max}}{2} B_k$ and collecting the two B_k contributions,

$$B_k + \frac{6\sqrt{k}}{\gamma_{\alpha}\nu}\Big(2\zeta_{\delta} + \frac{q_{\max}}{2} B_k\Big) = \Big(1 + \frac{3q_{\max}\sqrt{k}}{\gamma_{\alpha}\nu}\Big) B_k + \frac{12\sqrt{k}}{\gamma_{\alpha}\nu}\zeta_{\delta},$$

which is the second bound in Theorem 1. Hence,

$$\|\widehat{\Theta}_k - \Theta^{\star}\|_F \le \Big(1 + \frac{3q_{\max}\sqrt{k}}{\gamma_{\alpha}\nu}\Big) B_k + \frac{12\sqrt{k}}{\gamma_{\alpha}\nu}\zeta_{\delta}, \tag{31}$$

The bound on B_k in terms of $\varsigma_{>k}$ is Proposition 2, and the corresponding margin statement follows from Lemma 1. □

B.5 Rates and ranking consequences

B.5.1 Balanced-design rates and consistency

Corollary B.1 (Rate, sample complexity, consistency)**.** *Take a complete pair set with $N_{ij} \equiv \bar{N}$ and $q_{ij} \equiv \bar{q} = \binom{M}{2}^{-1}$, so that $\nu = q_{\max} = q_{\min} = \bar{q}$ and $n = \bar{N}\binom{M}{2}$. Then $\bar{q}$ cancels from equation 31 and, with probability at least $1-\delta$,*

$$\frac{\|\widehat{\Theta}_k - \Theta^\star\|_F}{M} \;\le\; \Big(1 + \frac{3\sqrt{k}}{\gamma_\alpha}\Big)\frac{B_k}{M} + \frac{C}{\gamma_\alpha}\sqrt{\frac{k\,M\log(M/\delta)}{n}} \tag{32}$$

for an absolute constant C. If the class contains the truth, $\operatorname{rank}\Theta^\star \le 2k$, then $B_k = 0$; per-entry accuracy ε is reached once $n \gtrsim kM\log M/(\gamma_\alpha^2\varepsilon^2)$, i.e. after $\widetilde{O}(k/\varepsilon^2)$ comparisons per agent, and $M^{-2}\mathbb{E}\|\widehat{\Theta}_k - \Theta^\star\|_F^2 \to 0$ as $\bar{N} \to \infty$. The same statements hold for the margin matrix after halving, by Lemma 1.

Proof. Step 1. Uniform weights give $q_{\max} = q_{\min} = \bar{q}$ and, by Definition 1, $\nu = \bar{q}$ exactly. From the definitions in Proposition 5 and $\sigma' \le \frac{1}{4}$,

$$d_\star = \max_i \sum_{j\neq i} \frac{\bar{q}^2\sigma'(\Theta^\star_{ij})}{\bar{N}} \le \frac{\bar{q}^2(M-1)}{4\bar{N}}, \qquad \mathsf{q}_\star = \frac{\bar{q}}{\bar{N}}.$$

Writing $\ell_\delta := \log(2M/\delta)$ and substituting into equation 30,

$$\zeta_\delta \;\le\; \sqrt{2\cdot\frac{\bar{q}^2(M-1)}{4\bar{N}}\,\ell_\delta} + \frac{2}{3}\cdot\frac{\bar{q}}{\bar{N}}\ell_\delta \;=\; \bar{q}\Big(\sqrt{\tfrac{(M-1)\ell_\delta}{2\bar{N}}} + \tfrac{2\ell_\delta}{3\bar{N}}\Big).$$

Step 2. Insert this and $\nu = q_{\max} = \bar{q}$ into the second bound of equation 5. The factor $\bar{q}$ appears once in the numerator (through ζ_δ) and once in the denominator (through ν), so it cancels:

$$\|\widehat{\Theta}_k - \Theta^\star\|_F \;\le\; \Big(1 + \tfrac{3\sqrt{k}}{\gamma_\alpha}\Big)B_k + \frac{12\sqrt{k}}{\gamma_\alpha}\Big(\sqrt{\tfrac{(M-1)\ell_\delta}{2\bar{N}}} + \tfrac{2\ell_\delta}{3\bar{N}}\Big).$$

Step 3. From $n = \bar{N}\binom{M}{2}$ we get $\bar{N} = 2n/(M(M-1))$, so

$$\sqrt{\frac{(M-1)\ell_\delta}{2\bar{N}}} = \sqrt{\frac{(M-1)\ell_\delta\,M(M-1)}{4n}} = \frac{(M-1)\sqrt{M\ell_\delta}}{2\sqrt{n}}.$$

The second term is dominated by the first whenever $\frac{2\ell_\delta}{3\bar{N}} \le \frac{(M-1)\sqrt{M\ell_\delta}}{2\sqrt{n}}$, which after substituting $\bar{N}$ reduces to $\frac{2}{3}\sqrt{M\ell_\delta/n} \le 1$, i.e. $n \ge M\ell_\delta$ — a condition weaker than requiring the comparison graph to be connected. Dividing Step 2 by M and absorbing constants gives equation 32.

Step 4 (sample complexity). Set $B_k = 0$ and require the right side of equation 32 to be at most ε:

$$\frac{C}{\gamma_\alpha}\sqrt{\frac{kM\log(M/\delta)}{n}} \le \varepsilon \iff n \ge \frac{C^2 kM\log(M/\delta)}{\gamma_\alpha^2\varepsilon^2}.$$

Each comparison involves two agents, so n comparisons provide $2n/M$ per agent, giving $\widetilde{O}(k/\varepsilon^2)$ per agent.

Step 5. Take $\delta = (M\bar{N})^{-1}$, so $\ell_\delta = \log(2M^2\bar{N})$, and let B_δ denote the right-hand side of the display in Step 2 with $B_k = 0$. On the good event $\|\widehat{\Theta}_k - \Theta^\star\|_F \le B_\delta$. Off it, (A1) still gives the deterministic bound $\|\widehat{\Theta}_k - \Theta^\star\|_{\max} \le 2\alpha$, hence by Lemma 2

$$\|\widehat{\Theta}_k - \Theta^\star\|_F^2 = 2\sum_{i<j}(\cdot)^2 \le 2\binom{M}{2}(2\alpha)^2 = 4\alpha^2 M(M-1) \le 4\alpha^2 M^2.$$

Splitting the expectation over the two events,

$$\frac{\mathbb{E}\|\widehat{\Theta}_k - \Theta^\star\|_F^2}{M^2} \;\le\; \frac{B_\delta^2}{M^2} + 4\alpha^2\delta \;\le\; \frac{C\,k\log(2M^2\bar{N})}{\gamma_\alpha^2\,M\bar{N}} + \frac{4\alpha^2}{M\bar{N}} \;\xrightarrow[\bar{N}\to\infty]{}\; 0,$$

where the middle step uses $(a+b)^2 \le 2a^2 + 2b^2$ on B_δ and absorbs the squared Bernstein tail via $\log(2M^2\bar{N}) \le M\bar{N}$, valid for $M \ge 3$. □

B.6 SUPPORTING RESULTS

B.6.1 STANDARD MATRIX AND CONCENTRATION FACTS

Lemma 11 (Real Schur form). *If $\Theta \in \mathbb{S}^K$ has $\operatorname{rank}(\Theta) = 2r \le 2k \le M$, there are $Q \in \mathbb{R}^{M\times 2k}$ with $Q^\top Q = I_{2k}$ and $\varsigma_1 \ge \cdots \ge \varsigma_k \ge 0$ such that $\Theta = Q\,\mathrm{blkdiag}(\varsigma_1 J, \ldots, \varsigma_k J)Q^\top$. Each ς_l is a singular value of multiplicity two, so*

$$\|\Theta\|_* = 2\sum_{l=1}^{k}\varsigma_l, \qquad \|\Theta\|_F^2 = 2\sum_{l=1}^{k}\varsigma_l^2. \tag{33}$$

Proof. Θ is normal matrix ($\Theta^\top\Theta = -\Theta^2 = \Theta\Theta^\top$), so the real Schur decomposition applies and yields an orthogonal $\widetilde{Q} \in \mathbb{R}^{M\times M}$ with $\Theta = \widetilde{Q}\,\mathrm{blkdiag}(\varsigma_1 J, \ldots, \varsigma_r J, 0, \ldots, 0)\widetilde{Q}^\top$ and $\varsigma_1 \ge \cdots \ge \varsigma_r > 0$. Keep the first $2r$ columns of $\widetilde{Q}$ and append $2(k-r)$ further columns (possible since $M \ge 2k$), setting $\varsigma_l := 0$ for $r < l \le k$. For equation 33, the block $\varsigma_l J$ has singular values ς_l, ς_l, so the singular values of Θ are each ς_l repeated twice, together with zeros; summing them gives the nuclear norm and summing their squares gives the squared Frobenius norm. □

Lemma 12 (Norm facts). (i) $\|X\|_* = \max_{\|B\|_{\mathrm{op}}\le 1}\langle X, B\rangle$, *hence* $|\langle X, Y\rangle| \le \|X\|_{\mathrm{op}}\|Y\|_*$; (ii) $\|X^\top Y\|_* \le \|X\|_F\|Y\|_F$; (iii) $\|BX\|_F \le \|B\|_{\mathrm{op}}\|X\|_F$; (iv) *if* $\operatorname{rank} X \le r$ *then* $\|X\|_* \le \sqrt{r}\,\|X\|_F$.

Proof. (i) First part is the duality between the Schatten-∞ and Schatten-1 norms. The second part follows by applying it to Y with $B = X/\|X\|_{\mathrm{op}}$.

(ii) Fix B with $\|B\|_{\mathrm{op}} \le 1$. Then

$$\langle X^\top Y, B\rangle = \mathrm{Tr}(Y^\top XB) = \langle X, YB^\top\rangle \le \|X\|_F\|YB^\top\|_F \le \|X\|_F\|Y\|_F\|B\|_{\mathrm{op}} \le \|X\|_F\|Y\|_F,$$

using Cauchy–Schwarz and then (iii). Maximizing the left side over B and applying (i) gives (ii).

(iii) Writing $x_1, \ldots, x_M$ for the columns of X, $\|BX\|_F^2 = \sum_l \|Bx_l\|_2^2 \le \|B\|_{\mathrm{op}}^2 \sum_l \|x_l\|_2^2$.

(iv) By Cauchy–Schwarz on the at most r nonzero singular values,

$$\|X\|_* = \sum_{l\le r}\varsigma_l(X) \le \sqrt{r}\Big(\sum_l \varsigma_l(X)^2\Big)^{1/2} = \sqrt{r}\,\|X\|_F.$$

□

Lemma 13 (Matrix Bernstein; (Tropp, 2015)Theorem 6.1.1). *Let $X_1, \ldots, X_T$ be independent, mean-zero, Hermitian $M \times M$ random matrices with $\|X_t\|_{\mathrm{op}} \le R$ almost surely, and let $v = \|\sum_t \mathbb{E}X_t^2\|_{\mathrm{op}}$. For any $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\|\textstyle\sum_t X_t\|_{\mathrm{op}} \le \sqrt{2v\log(2M/\delta)} + \frac{2}{3}R\log(2M/\delta).$$

Lemma 14 (Hoeffding; Boucheron et al., 2013, Section 2.3). *A centered random variable supported on an interval of length d is sub-Gaussian with proxy $d/2$. Consequently a centered $\mathrm{Bin}(N,p)$ variable, being a sum of N independent centered Bernoullis each supported on an interval of length 1, has proxy $\sqrt{N}/2$.*

B.6.2 INVARIANCE UNDER DESIGN-MATCHED REWEIGHTING

Proposition 6 (Design-matched weights recover the raw likelihood). *Let $\widehat{\mathcal{L}}^{\mathrm{tot}}(\Theta) = \sum_{\mathcal{E}}[N_{ij}\psi(\Theta_{ij}) - W_{ij}\Theta_{ij}]$ be the raw total negative log-likelihood, with score Z^{tot}, restricted eigenvalue ν^{tot} defined by $\sum_{\mathcal{E}} N_{ij}\Upsilon_{ij}^2 \ge \frac{\nu^{\mathrm{tot}}}{2}\|\Upsilon\|_F^2$, and weighted degree $d_\star^{\mathrm{tot}} = \max_i \sum_j N_{ij}\sigma'(\Theta^\star_{ij})$. If the weights are design-matched, $q_{ij} = N_{ij}/n$, then*

$$\widehat{\mathcal{L}} = \tfrac{1}{n}\widehat{\mathcal{L}}^{\mathrm{tot}}, \qquad Z = \tfrac{1}{n}Z^{\mathrm{tot}}, \qquad \nu = \tfrac{1}{n}\nu^{\mathrm{tot}}, \qquad d_\star = \tfrac{1}{n^2}d_\star^{\mathrm{tot}},$$

so that $\sqrt{d_\star}/\nu$ is unchanged and equation 22 is identical under the two normalizations.

Proof. Step 1. Substituting $q_{ij} = N_{ij}/n$ and $\widehat{P}_{ij} = W_{ij}/N_{ij}$,

$$\widehat{\mathcal{L}}(\Theta) = \sum_{\mathcal{E}} \frac{N_{ij}}{n}\Big[\psi(\Theta_{ij}) - \frac{W_{ij}}{N_{ij}}\Theta_{ij}\Big] = \frac{1}{n}\sum_{\mathcal{E}}\big[N_{ij}\psi(\Theta_{ij}) - W_{ij}\Theta_{ij}\big] = \frac{1}{n}\widehat{\mathcal{L}}^{\mathrm{tot}}(\Theta).$$

Step 2. $Z_{ij} = q_{ij}(P^\star_{ij} - \widehat{P}_{ij}) = \frac{1}{n}(N_{ij}P^\star_{ij} - W_{ij}) = \frac{1}{n}Z^{\mathrm{tot}}_{ij}$.

Step 3 (restricted eigenvalue). $\sum_{\mathcal{E}} q_{ij}\Upsilon^2_{ij} = \frac{1}{n}\sum_{\mathcal{E}} N_{ij}\Upsilon^2_{ij}$, so the best constant in equation 21 scales by $1/n$.

Step 4 (weighted degree).

$$d_\star = \max_i \sum_j \frac{(N_{ij}/n)^2\sigma'(\Theta^\star_{ij})}{N_{ij}} = \frac{1}{n^2}\max_i \sum_j N_{ij}\sigma'(\Theta^\star_{ij}) = \frac{1}{n^2}d^{\mathrm{tot}}_\star.$$

Step 5 (the ratio is invariant). By Proposition 5 the admissible λ scales like $\sqrt{d_\star}$, hence by $1/n$; and ν scales by $1/n$ as well. The bound equation 22 depends on these only through λ/ν, which is therefore unchanged. □

Therefore, reweighting changes only the estimand. Changing relative pair weights can change both the population target and the error bound. For a fixed observed pair set, choosing reference weights in advance defines a population target independent of comparison frequencies. This motivates the equal weights over pairs in Section 5.

C Readout definitions and guarantees

Throughout, A is antisymmetric, $M \geq 2$, $\tau > 0$, and logarithms are natural. Unless a different bound is stated, $A \in [-1,1]^{M\times M}$. We write $a_{\max} = \|A\|_{\max}$, $p = p^\tau(A)$, $s = Ap$, and $H(p) = -\sum_i p_i \log p_i$.

C.1 Existence, uniqueness, and the temperature branch

The continuous map $x \mapsto \mathrm{softmax}(Ax/\tau)$ maps the compact simplex into itself, so Brouwer's theorem gives a fixed point. Every fixed point has full support. If p and q are two fixed points, subtracting their log-softmax equations and pairing with $p - q$ gives

$$\tau\big[\mathrm{KL}(p\|q) + \mathrm{KL}(q\|p)\big] = (p-q)^\top A(p-q) = 0.$$

Thus $p = q$. For any sequence $\tau_n \to \tau_0 > 0$, compactness gives a convergent subsequence of $p^{\tau_n}(A)$; continuity of the fixed-point equation makes its limit the unique $p^{\tau_0}(A)$. Every convergent subsequence has this limit, proving continuity in $\tau > 0$. The same argument gives qualitative continuity in A at fixed positive temperature; the quantitative bound is proved below.

C.2 Condorcet concentration and ordering

Theorem 2(i) (Approximate Condorcet consistency). *Let c be a strict Condorcet winner, with $g_{\mathrm{CW}} = \min_{j\neq c} A_{cj} > 0$ and $a_{\max} = \|A\|_{\max}$. Then QRE ranks c strictly first whenever $\tau < g^2_{\mathrm{CW}}/(a_{\max}\log M)$.*

Proof. **Step 1 (bound the largest score).** Let $Z = \sum_k e^{s_k/\tau}$. The fixed-point equation gives $s_i = \tau\log p_i + \tau\log Z$. Multiplying by p_i, summing, and using antisymmetry,

$$0 = p^\top Ap = \sum_i p_i s_i = -\tau H(p) + \tau\log Z, \qquad \log Z = H(p).$$

The maximum score is at least its p-weighted average, which is zero. Also, $H(p) = \log M - \mathrm{KL}(p\|\mathbf{1}/M) \leq \log M$ and $\log\|p\|_\infty \leq 0$. Hence

$$0 \leq \max_i s_i = \tau\big(H(p) + \log\|p\|_\infty\big) \leq \tau\log M. \tag{34}$$

Step 2 (bound the mass outside the winner). Set $v = 1 - p_c = \sum_{k \neq c} p_k$. Since $A_{cc} = 0$ and $A_{ck} \geq g_{\mathrm{CW}}$ for $k \neq c$,

$$g_{\mathrm{CW}} v \leq \sum_{k \neq c} A_{ck} p_k = s_c \leq \max_i s_i \leq \tau \log M.$$

Dividing by $g_{\mathrm{CW}} > 0$ gives

$$v \leq \frac{\tau \log M}{g_{\mathrm{CW}}}. \tag{35}$$

Step 3 (compare the winner with each rival). For $j \neq c$, antisymmetry gives $A_{jc} \leq -g_{\mathrm{CW}}$. Using $A_{jj} = 0$ and $A_{jk} \leq a_{\max}$,

$$s_j = A_{jc} p_c + \sum_{k \notin \{c,j\}} A_{jk} p_k \leq -g_{\mathrm{CW}} p_c + a_{\max}(v - p_j).$$

Subtract this bound from $s_c \geq g_{\mathrm{CW}} v$ and use $v + p_c = 1$:

$$\begin{aligned} s_c - s_j &\geq g_{\mathrm{CW}}(v + p_c) - a_{\max}(v - p_j) \\ &= g_{\mathrm{CW}} - a_{\max} v + a_{\max} p_j \\ &\geq g_{\mathrm{CW}} - a_{\max} v \geq g_{\mathrm{CW}} - \frac{a_{\max} \tau \log M}{g_{\mathrm{CW}}}. \end{aligned} \tag{36}$$

The last expression is positive when $\tau < g_{\mathrm{CW}}^2 / (a_{\max} \log M)$, so $s_c > s_j$ for every $j \neq c$. □

C.3 Insertion: direct and redistribution effects

Theorem 2(ii) (Approximate independence of irrelevant alternatives). *Let $A' \in [-1,1]^{(M+1)\times(M+1)}$ be antisymmetric, with $A'_{ij} = A_{ij}$ for $i, j \in [M]$, and let $a_i = A'_{iu}$. At fixed $\tau > 0$, write $s' = s^\tau(A')$ and $\mu = p_u^\tau(A')$. Define*

$$\beta(A) = \sup_{\mathbf{1}^\top x = 0,\, x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_1} \leq a_{\max} \leq 1.$$

For all $i, j \in [M]$,

$$\left|(s'_i - s'_j) - (s_i - s_j)\right| \leq \mu \left(|(a_i - a_j) - (s_i - s_j)| + \frac{2\beta(A)}{\tau} \|a - s\|_\infty \right) =: \varepsilon_{\mathrm{IIA}}^{ij}.$$

If $\Gamma = \max_{k \in [M]} s'_k - s'_u \geq 0$, then $\mu \leq (1 + e^{\Gamma/\tau})^{-1} \leq e^{-\Gamma/\tau}$.

Proof. **Step 1 (condition on the incumbent models).** Write $p' = p^\tau(A')$ and $q_i = p'_i/(1-\mu)$ for $i \in [M]$. Full support gives $0 < \mu < 1$, so q is a probability vector with positive entries. The expanded scores and the conditional softmax satisfy

$$s'_{[M]} = (1-\mu) A q + \mu a, \qquad q_i = \frac{p'_i}{\sum_{k \in [M]} p'_k} = \frac{e^{s'_i/\tau}}{\sum_{k \in [M]} e^{s'_k/\tau}}.$$

Step 2 (separate the two insertion effects). Since $s = Ap$,

$$s'_{[M]} - s = (1-\mu) A q + \mu a - A p = (1-\mu) A (q - p) + \mu (a - s).$$

Taking the difference of coordinates i and j gives

$$\begin{aligned} (s'_i - s'_j) - (s_i - s_j) = (1-\mu)\big([A(q-p)]_i - [A(q-p)]_j\big) \\ + \mu\big((a_i - a_j) - (s_i - s_j)\big). \end{aligned} \tag{37}$$

The first term is the redistribution of incumbent weights; the second is the entrant's direct effect.

Step 3 (bound the redistribution of weights). Let $h = q - p$, $Z_p = \sum_{k \in [M]} e^{s_k/\tau}$, and $Z_q = \sum_{k \in [M]} e^{s'_k/\tau}$. Take logarithms of $q_i = e^{s'_i/\tau}/Z_q$ and $p_i = e^{s_i/\tau}/Z_p$, then subtract and use Step 2:

$$\tau(\log q - \log p) = (1-\mu) A h + \mu (a - s) - \tau \log(Z_q/Z_p) \mathbf{1}.$$

Take the inner product with h. Since $\mathbf{1}^\top h = 0$ and $h^\top A h = 0$,

$$\tau\big[\mathrm{KL}(q\|p) + \mathrm{KL}(p\|q)\big] = \tau \sum_i (q_i - p_i)\log(q_i/p_i)$$
$$= \mu h^\top (a - s) \le \mu \|a - s\|_\infty \|h\|_1.$$

Pinsker's inequality gives $\mathrm{KL}(q\|p) \ge \frac{1}{2}\|h\|_1^2$ and $\mathrm{KL}(p\|q) \ge \frac{1}{2}\|h\|_1^2$. Thus $\tau\|h\|_1^2 \le \mu\|a - s\|_\infty \|h\|_1$. If $h = 0$ the desired bound is immediate; otherwise divide by $\tau\|h\|_1$ to obtain

$$\|q - p\|_1 \le \frac{\mu}{\tau}\|a - s\|_\infty.$$

Step 4 (bound each score-gap change). Because $q - p$ has zero sum, $\|A(q-p)\|_\infty \le \beta(A)\|q-p\|_1$. Apply the triangle inequality to equation 37, then use Step 3:

$$|(s'_i - s'_j) - (s_i - s_j)| \le 2(1-\mu)\beta(A)\|q - p\|_1 + \mu|(a_i - a_j) - (s_i - s_j)|$$
$$\le \mu\left(|(a_i - a_j) - (s_i - s_j)| + \frac{2(1-\mu)\beta(A)}{\tau}\|a - s\|_\infty\right)$$
$$\le \varepsilon_{\mathrm{IIA}}^{ij}.$$

Here the factor 2 bounds the difference of two coordinates, and the last line uses $1 - \mu \le 1$.

Step 5 (bound the entrant's mass). Divide the expanded softmax denominator by $e^{s'_u/\tau}$:

$$\mu = \frac{1}{1 + \sum_{k\in[M]} e^{(s'_k - s'_u)/\tau}} \le \frac{1}{1 + e^{\Gamma/\tau}} \le e^{-\Gamma/\tau} \quad (\Gamma \ge 0).$$

The first inequality retains only the largest incumbent term in the sum. □

Strict-loser insertion. For a fixed strict loser, set $d = \min_i a_i > 0$ and $\bar{a} = q^\top a \ge d$. Antisymmetry gives

$$s'_u = -(1-\mu)\bar{a}, \qquad \sum_i q_i s'_i = (1-\mu)q^\top A q + \mu q^\top a = \mu\bar{a}.$$

The largest incumbent score is at least this weighted average, so

$$\Gamma = \max_{i\in[M]} s'_i - s'_u \ge \mu\bar{a} + (1-\mu)\bar{a} = \bar{a} \ge d.$$

Since $|(a_i - s_i) - (a_j - s_j)| \le 2\|a - s\|_\infty$, $\beta(A) \le 1$, and $\|a - s\|_\infty \le 2$, the insertion bound satisfies

$$\varepsilon_{\mathrm{IIA}}^{ij} \le 2\mu(1 + \tau^{-1})\|a - s\|_\infty \le 4(1 + \tau^{-1})e^{-d/\tau} \longrightarrow 0 \quad \text{as } \tau \downarrow 0.$$

C.4 Matrix perturbations and statistical coverage

Theorem 3 (Stability of QRE ranking). *For antisymmetric $A, B \in [-1, 1]^{M\times M}$, fixed $\tau > 0$, and $\varepsilon = \|A - B\|_{\max}$,*

$$\|p^\tau(A) - p^\tau(B)\|_1 \le \frac{\varepsilon}{\tau}, \qquad \|s^\tau(A) - s^\tau(B)\|_\infty \le (1 + \tau^{-1})\varepsilon.$$

Proof. **Step 1 (subtract the equilibrium equations).** Write $p = p^\tau(A)$, $q = p^\tau(B)$, and $h = p - q$. Both distributions have full support. With $Z_A = \sum_i e^{(Ap)_i/\tau}$ and $Z_B = \sum_i e^{(Bq)_i/\tau}$,

$$\tau(\log p - \log q) = Ap - Bq - \tau\log(Z_A/Z_B)\mathbf{1}.$$

Take the inner product with h. The normalization term vanishes because $\mathbf{1}^\top h = 0$. Also, $Ap - Bq = Ah + (A - B)q$ and $h^\top A h = 0$. Thus

$$\tau\big[\mathrm{KL}(p\|q) + \mathrm{KL}(q\|p)\big] = \tau \sum_i (p_i - q_i)\log(p_i/q_i)$$
$$= h^\top(Ap - Bq) = h^\top(A - B)q.$$

Step 2 (bound the probability change). Since $q_j \geq 0$ and $\sum_j q_j = 1$,

$$\|(A-B)q\|_\infty \leq \max_i \sum_j |A_{ij} - B_{ij}| q_j \leq \varepsilon \sum_j q_j = \varepsilon.$$

Pinsker's inequality bounds each KL term below by $\frac{1}{2}\|h\|_1^2$. Combining Step 1 with norm duality gives

$$\tau \|h\|_1^2 \leq h^\top (A-B)q \leq \|h\|_1 \|(A-B)q\|_\infty \leq \varepsilon \|h\|_1.$$

For $h \neq 0$, divide by $\tau\|h\|_1$; for $h = 0$, the bound is immediate. Thus

$$\|p - q\|_1 \leq \frac{\varepsilon}{\tau}.$$

Step 3 (transfer the bound to scores). The unit payoff bound gives

$$\|A(p-q)\|_\infty \leq \max_i \sum_j |A_{ij}|\,|p_j - q_j| \leq \|p-q\|_1.$$

Using $Ap - Bq = A(p-q) + (A-B)q$ and Step 2,

$$\begin{aligned} \|Ap - Bq\|_\infty &\leq \|A(p-q)\|_\infty + \|(A-B)q\|_\infty \\ &\leq \|p-q\|_1 + \varepsilon \leq (1+\tau^{-1})\varepsilon. \end{aligned}$$

□

Population score ordering. Under the main-text estimation theorem's assumptions, let $\widehat{\Theta}$ be its estimator and $R_\Theta(\delta)$ its Frobenius error bound. Set $\widehat{A}_{ij} = \tanh(\widehat{\Theta}_{ij}/2)$, $A^\star_{ij} = \tanh(\Theta^\star_{ij}/2)$, and $\widehat{s} = s^\tau(\widehat{A})$ at fixed $\tau > 0$. For $\delta \in (0,1)$, with probability at least $1-\delta$, simultaneously for all i, j,

$$\widehat{s}_i - \widehat{s}_j > (1+\tau^{-1}) R_\Theta(\delta) \quad \Longrightarrow \quad s_i^\tau(A^\star) > s_j^\tau(A^\star). \tag{38}$$

Proof. **Step 1 (convert log-odds error to margin error).** The estimation theorem gives an event $\mathcal{G}_\delta$ of probability at least $1-\delta$ on which $\|\widehat{\Theta} - \Theta^\star\|_F \leq R_\Theta(\delta)$. For $f(x) = \tanh(x/2)$, $|f'(x)| = \frac{1}{2}\operatorname{sech}^2(x/2) \leq \frac{1}{2}$. The mean value theorem therefore gives $|\widehat{A}_{ij} - A^\star_{ij}| \leq \frac{1}{2}|\widehat{\Theta}_{ij} - \Theta^\star_{ij}|$. On $\mathcal{G}_\delta$,

$$\|\widehat{A} - A^\star\|_{\max} \leq \tfrac{1}{2}\|\widehat{\Theta} - \Theta^\star\|_{\max} \leq \tfrac{1}{2}\|\widehat{\Theta} - \Theta^\star\|_F \leq \tfrac{1}{2} R_\Theta(\delta).$$

Step 2 (apply QRE stability). Both margin matrices are antisymmetric with entries in $[-1, 1]$. Apply Proposition 3 with $A = A^\star$ and $B = \widehat{A}$:

$$\|\widehat{s} - s^\tau(A^\star)\|_\infty \leq (1+\tau^{-1})\|\widehat{A} - A^\star\|_{\max} \leq \tfrac{1}{2}(1+\tau^{-1}) R_\Theta(\delta) =: e_\delta.$$

Step 3 (subtract the errors in both scores). For every pair on $\mathcal{G}_\delta$, the triangle inequality gives

$$\begin{aligned} |(\widehat{s}_i - \widehat{s}_j) - (s_i^\tau(A^\star) - s_j^\tau(A^\star))| &\leq |\widehat{s}_i - s_i^\tau(A^\star)| + |\widehat{s}_j - s_j^\tau(A^\star)| \\ &\leq 2e_\delta = (1+\tau^{-1}) R_\Theta(\delta). \end{aligned}$$

Thus $s_i^\tau(A^\star) - s_j^\tau(A^\star) \geq \widehat{s}_i - \widehat{s}_j - 2e_\delta > 0$ whenever the displayed condition holds. Every pairwise implication holds on the same event $\mathcal{G}_\delta$, giving the stated simultaneous probability. □

C.5 Exact duplication and the softmax offset

Fix $\tau > 0$. Replace model c by $\kappa \geq 1$ exact copies $c_1, \ldots, c_\kappa$, where κ is an integer. Each copy has model c's comparisons against every unchanged model, and margins between copies are zero. All other comparisons stay fixed. Let A' be the expanded matrix, $p' = p^\tau(A')$, and $s' = A'p'$. Aggregate the copies into $v \in \Delta_M$ by setting

$$v_i = p'_i \quad (i \neq c), \qquad v_c = \sum_{\ell=1}^{\kappa} p'_{c_\ell}.$$

Then the aggregated weights satisfy

$$v = \operatorname{softmax}\left(\frac{Av + \tau(\log\kappa) e_c}{\tau}\right), \tag{39}$$

where e_c is the c-th coordinate vector.

Proof. **Step 1 (express the expanded scores using v).** For each unchanged model $i \neq c$ and each copy c_ℓ,

$$s'_i = \sum_{j \neq c} A_{ij} p'_j + A_{ic} \sum_{\ell=1}^{\kappa} p'_{c_\ell} = (Av)_i,$$
$$s'_{c_\ell} = \sum_{j \neq c} A_{cj} p'_j = (Av)_c.$$

Here the copy-to-copy terms vanish, and $A_{cc} = 0$. Thus all copies have the same score and, by the QRE softmax equation, the same individual probability.

Step 2 (sum the copies' softmax probabilities). The expanded softmax denominator is

$$Z' = \sum_{j \neq c} e^{(Av)_j/\tau} + \kappa e^{(Av)_c/\tau}.$$

Hence

$$v_i = \frac{e^{(Av)_i/\tau}}{Z'} \quad (i \neq c), \qquad v_c = \sum_{\ell=1}^{\kappa} \frac{e^{(Av)_c/\tau}}{Z'} = \frac{\kappa e^{(Av)_c/\tau}}{Z'}.$$

Step 3 (write the multiplicity as an offset). Since

$$\kappa e^{(Av)_c/\tau} = e^{((Av)_c + \tau \log \kappa)/\tau},$$

substituting this expression in the numerator and denominator gives equation 39. □

The offset $\tau(\log \kappa) e_c$ accounts for the number of identical copies in softmax. The ranking scores are Av, evaluated at the new aggregated weights v; the offset is not added to those scores. At fixed payoff scores, duplication multiplies the family's unnormalized softmax weight by κ. This changes the shared opponent distribution and can alter the incumbent ranking. For any $i \neq c$,

$$\frac{v_c}{v_i} = \kappa \exp\left(\frac{(Av)_c - (Av)_i}{\tau}\right).$$

The direct cloning bias is small relative to a payoff gap $g > 0$ when $\tau \log \kappa \ll g$, and becomes comparable when $\tau \log \kappa \gtrsim g$. It grows linearly with τ and logarithmically with the number of copies. Thus QRE only provide approximate clone consistency.

C.6 Baseline readouts

The three baselines take the same complete margin matrix A as input. Write $P_{ij} = (1 + A_{ij})/2$ and $\sigma(x) = (1 + e^{-x})^{-1}$. BT and SCO weight all model pairs equally.

BT. BT fits ratings $r \in \mathbb{R}^M$, with predicted win probabilities $\sigma(r_i - r_j)$ (Bradley & Terry, 1952). Its cross-entropy fit is

$$r^{\mathrm{BT}}(A) \in \arg\min_{\mathbf{1}^\top r = 0} \sum_{i<j} \left[\log(1 + e^{r_i - r_j}) - P_{ij}(r_i - r_j)\right]. \tag{40}$$

Centering fixes the common additive shift. We assume a finite optimum and rank models by decreasing r_i^{BT}.

SCO. SCO fits scalar scores by smoothing pairwise ranking disagreement (Lanctot et al., 2025). For score bound $B > 0$ and smoothing parameter $\eta > 0$, define

$$t_\eta^{\mathrm{SCO}}(A) \in \arg\min_{t \in [-B,B]^M} L_{\mathrm{SCO}}(t; A),$$
$$L_{\mathrm{SCO}}(t; A) = \sum_{i<j} \left[P_{ij}\sigma\left(\frac{t_j - t_i}{\eta}\right) + P_{ji}\sigma\left(\frac{t_i - t_j}{\eta}\right)\right]. \tag{41}$$

Models are ranked by decreasing score. Equation equation 41 defines global minimizers; the experiments report the output of the numerical solver specified in Appendix D.2.

For a strict order π, with π_a ahead of π_b when $a < b$, the Kemeny disagreement is $L_{\mathrm{K}}(\pi; P) = \sum_{a<b} P_{\pi_b, \pi_a}$. For any fixed score vector with distinct entries and induced order π, $L_{\mathrm{SCO}}(t; A) \to L_{\mathrm{K}}(\pi; P)$ as $\eta \downarrow 0$.

MaxL. MaxL selects a lottery over models that no opponent can beat in expected margin (Fishburn, 1984). We use the correspondence $\mathcal{M}(A)$ defined in the main text. Selecting $p \in \mathcal{M}(A)$ returns a probability distribution, whose support is $\{i : p_i > 0\}$.

D EXPERIMENTAL PROTOCOLS AND ADDITIONAL RESULTS

D.1 DATASETS, RESAMPLING, AND SCORE SEPARATION

Datasets and comparison weights. Arena uses the `lmarena-ai/arena-human-preference-140k`. Removing ten self-matches leaves 135,624 comparisons among 53 models. Arena28 is the lexicographically first maximum-cardinality clique whose pairs each have at least 100 comparisons. It contains 28 models, 378 pairs, and 71,866 votes; selection uses counts alone. Each pair's win rate is its mean outcome, with ties and both-bad records contributing half credit. The Arena48 pool used for prediction and completion is defined in Appendix D.5.

Our Datapoint Image Bench pool contains 30 models, 435 pairs, and 500 prompts, with 50 prompts in each of ten categories. We compute win rates within category–pair cells and average the ten category rates equally. Unannotated prompt–pair combinations contribute no observation. BT/SCO fitting weights model pairs equally in both datasets. Model lists and snapshot fingerprints are provided with the experimental artifacts.

Resampling. For Table 2, all readouts share 2,000 bootstrap samples within each dataset. Arena28 resamples 71,866 individual votes with replacement, implemented as multinomial draws over pair–outcome counts with outcomes in $\{0, \frac{1}{2}, 1\}$. Both pair counts and outcomes vary; a tie remains one half-credit observation. Image resamples 50 whole prompts with replacement within each category, retaining all associated annotations. Each draw recomputes the win rates using the same weighting rules and retains the original model pool.

Figure 1(c) instead uses the constructed four-model tournament before insertion. Each of its six pairs has 200 votes and is resampled independently with its count fixed, for 2,000 replicates; QRE uses $\tau = .075$.

Score intervals and pair separation. For each replicate, BT ratings are centered to have zero mean, SCO uses its scores in $[0, 100]$, and QRE uses uncentered payoff scores $s = Ap^\tau$ at the reported temperature. These normalization conventions are fixed across resamples because marginal score intervals depend on the score origin. Let $s_i^{(b)}$ be model i's score on replicate b. Its pointwise 95% percentile interval is

$$[L_i, U_i] = \left[Q_{.025}\{s_i^{(b)}\}_{b=1}^{2,000},\ Q_{.975}\{s_i^{(b)}\}_{b=1}^{2,000}\right],$$

using linear interpolation between order statistics. A pair is CI-separated when $L_i > U_j$ or $L_j > U_i$; touching endpoints do not count. This rule compares individual score intervals. Overlap leaves the pair unresolved by this rule and does not establish a tie or imply that an interval for the paired score difference contains zero.

Leader rank spread. For model i, the score intervals place j above i when $L_j > U_i$ and below i when $U_j < L_i$. The resulting rank band is

$$\left[1 + \#\{j \neq i : L_j > U_i\},\ M - \#\{j \neq i : U_j < L_i\}\right].$$

We apply this rule to the full-data leader to obtain its reported rank spread. These bands summarize the ordering constraints from pointwise score intervals; they are neither percentile intervals of resampled ranks nor simultaneous rank confidence sets. The summaries condition on the model pool, comparison weights, and temperature. SCO results additionally condition on the numerical solver. These empirical intervals do not instantiate the finite-sample matrix-error guarantee.

Set-valued scalar reports retain tied maxima. For Table 2, MaxL reports the unordered support of a selected maximal lottery, using $p_i > 10^{-7}$. Exact support recurrence differs from retention of the original leader.

D.2 CONTROLLED CYCLIC PERTURBATION

Perturbation. Figure 2(b) varies three pairwise win rates in the full-data Image tournament. Let S, G, and M denote Seedream 5.0 Pro, GPT Image 2, and MAI-Image-2.5, respectively, and let P be the original win-rate matrix. We form $P(\delta)$ by adding δ to P_{SG}, P_{GM}, and P_{MS}, decreasing their reverse entries by the same amount, and fixing all other pairs. Each affected model gains against one opponent and loses against another, so every row sum is preserved. We vary δ over $[0, \min\{1 - P_{SG}, 1 - P_{GM}, 1 - P_{MS}\}]$, the range that keeps all entries in $[0, 1]$. This raises Seedream's direct win rate against GPT from 48.70% to 95.11%.

For Arena28, we apply the same construction along ChatGPT-4o $\rightarrow$ Gemini 2.5 Pro $\rightarrow$ Gemini 2.5 Flash Lite $\rightarrow$ ChatGPT-4o. Over the full legal range, ChatGPT-4o's direct win rate against Gemini 2.5 Pro increases from 45.50% to 80.79%, while BT predicts 44.50% throughout. Figure 4 shows both datasets over their full perturbation ranges.

BT invariance. For the complete, equally weighted BT likelihood, this perturbation leaves the fitting objective unchanged. Writing $E = P(\delta) - P$, the loss change is proportional to $-r^\top E\mathbf{1} = 0$ for every rating vector r. Thus the centered BT ratings and all fitted pairwise probabilities remain fixed along the path; BT predicts a 49.42% win rate for Seedream against GPT throughout.

Readout evaluation. We recompute the readouts on each perturbed matrix, with QRE temperature $\tau = .075$. Immediately below and above $P_{SG}(\delta) = 1/2$, MaxL selects GPT and Seedream, respectively; the figures mark these one-sided limits. For Figure 2(b), SCO uses 10^7 stochastic gradient updates per input, learning rate 1, $\eta = 1$, batch size 32, and float32 scores initialized at 50 and clipped to $[0, 100]$. We run SCO ten times per input. Lines connect evaluated inputs; the SCO traces reflect optimizer variation on fixed matrices. These are controlled matrix perturbations, with no sampling intervals.

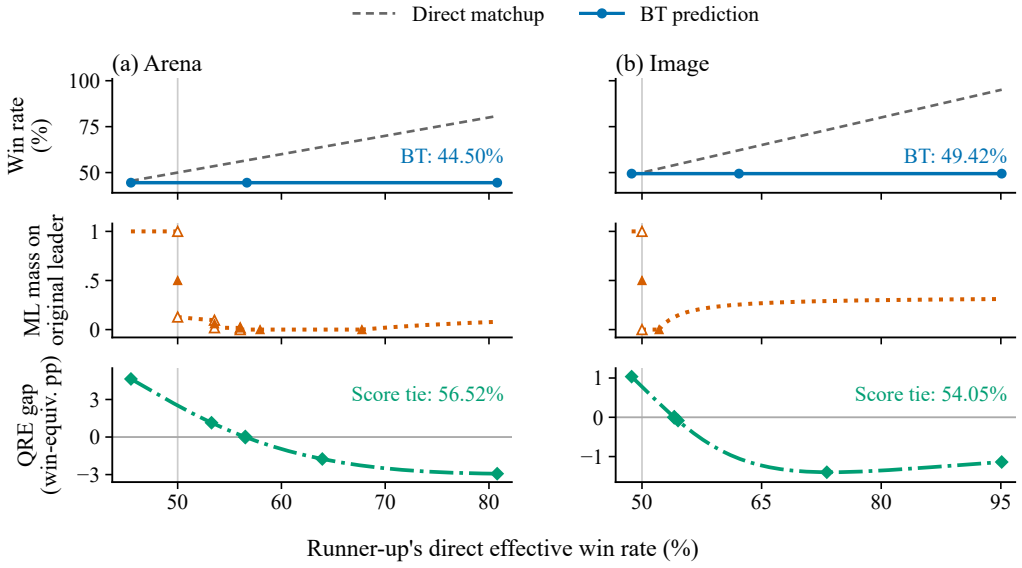


Figure 4: **BT remains fixed across the full cyclic perturbation paths.** Columns track ChatGPT-4o against Gemini 2.5 Pro (Arena28) and Seedream 5.0 Pro against GPT Image 2 (Image). Rows show direct and BT-fitted win probabilities, MaxL probability on the original leader c, and the QRE gap $50(s_c - s_d)$ against challenger d at $\tau = .075$ (win-equivalent percentage points). Open triangles mark MaxL one-sided limits; filled triangles mark selected boundary lotteries. MaxL lines connect evaluated points within support regions. These are deterministic matrix perturbations.

D.3 WEAK-ALTERNATIVE CHANGES

Image insertion. For Figure 2(a), we add a synthetic model w to the 30-model Image tournament, with $P_{iw} \in [.75, 1]$ and $P_{wi} = 1 - P_{iw}$ for every incumbent i. All 435 comparisons among incumbents remain fixed. Subject to changing the BT leader, the search prioritizes the number of incumbent rank positions changed, then the number of reversed pairs, then score separation. We fix the resulting entrant before evaluating SCO, MaxL, and QRE.

Arena deletion. We remove Claude 3.5 Haiku and then Amazon Nova Pro from Arena28. Each is a strict loser against every remaining model when removed. All 325 comparisons among the 26 survivors remain unchanged, and we recompute the readouts on the reduced tournament.

Ranking comparisons. We restrict both rankings to the common model set: the 30 Image incumbents or 26 Arena survivors. We then count models whose positions within that set change. SCO comparisons pair the same ten seeds before and after the intervention; Figure 2(a) reports the mean and range of changed positions. MaxL reports its unordered support. QRE uses $\tau = .075$ for the main comparison. For the temperature comparison, we hold the original and expanded Image matrices fixed and recompute both QRE rankings at each $\tau \in \{.01, .075, .5\}$, obtaining 0, 0, and 4 changed incumbent positions, respectively.

Temperature dependence. Figure 5 extends the comparison over $\tau \in [.01, 5]$. Panel (a) uses the same 2,000 bootstrap matrices per dataset at every temperature, with the resampling rules in Appendix D.1. Panel (b) holds the original and expanded Image matrices fixed. Mean Kendall distance decreases over the low-temperature range and then largely levels off. For the fixed Image entrant, increasing temperature eventually changes QRE's leader, although the original pool still ranks GPT Image 2 first. These results describe the selected inputs; they do not identify an optimal temperature.

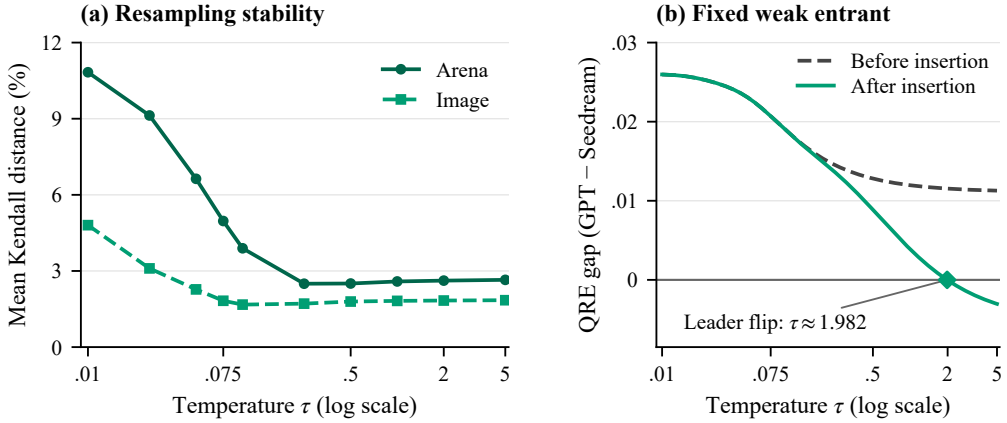


Figure 5: **Higher temperature can stabilize resampled rankings while allowing a weak entrant to change the leader.** (a) Mean normalized Kendall distance from the full-data ranking at the same temperature, reported as a percentage over 2,000 resamples. (b) The QRE score gap between GPT Image 2 and Seedream 5.0 Pro before and after the fixed Image insertion. After insertion, Seedream overtakes GPT near $\tau = 1.982$; the original pool retains GPT as leader throughout the displayed range.

D.4 MATRIX RECOVERY

Population and samples. For Figure 3(a), we treat the empirical distribution of the 71,866 Arena28 votes as the simulation population. Its pairwise mean outcomes, with half credit for ties, define $P^\star$ and the reference logits $\Theta^\star_{ij} = \text{logit}(P^\star_{ij})$. For each budget $n = 15{,}000 \cdot 2^j$, $j = 0, \ldots, 7$,

Table 5: Held-out prediction and fitted structure on Image Bench. NLL averages five outer prompt folds; cyclic fractions use a separate full-data fit.

Estimator	Held-out NLL ↓	Cyclic fraction (%)
BT	.683720	.00
DISC-1	.683676	.63
DISC-2	.683685	.67
DISC-4	.683683	.70

we draw 2,000 independent samples of n votes with replacement. This samples both model pairs and outcomes according to their observed joint frequencies. All estimators share each sample; different budgets use independent random streams. Every pair is observed in every generated sample.

Estimation. We fit DISC-k for $k \in \{1, 2, 4, 8\}$ using the objective in equation 3, with fixed pair weights $q_{ij} = 1/378$, logit bound $\alpha = 4$, and $\lambda = 0$ to isolate capacity and sampling effects. Each fit uses three initializations (spectral, perturbed spectral, and random), retaining the feasible solution with the lowest training loss. We use L-BFGS for interior fits and SLSQP with explicit logit constraints for endpoint win rates or out-of-bounds candidates. The empirical baseline is the pairwise logistic maximum-likelihood estimate under the same bound, $\widehat{\Theta}^{\mathrm{emp}}_{ij} = \mathrm{clip}(\mathrm{logit}\,\widehat{p}_{ij}, -4, 4)$. All fits use the sample win rates without pseudocounts.

Error summaries. For each sample, we compute $\|\widehat{\Theta} - \Theta^\star\|_F/M$ with $M = 28$, the root mean squared error over the matrix entries. The curves show the median, and the bands show the fifth and ninety-fifth percentiles across the 2,000 samples at each budget, using linear interpolation. These bands describe sampling variation in recovery error under the fixed empirical population.

D.5 Held-out prediction

Data splits. Arena48 retains models with at least 1,000 votes in the original 53-model pool, using counts alone. It contains 133,536 votes, 1,099 observed pairs, and 29 missing pairs. All estimators share a fixed split of 227 test pairs and 872 training pairs; holding out a pair removes all its votes. Image uses five outer folds of whole prompts, stratified by category, keeping each prompt's annotations together. Both datasets use three inner folds within each outer training set for regularization selection. Arena48 therefore evaluates prediction on unseen pairs, while Image evaluates new prompts for existing pairs.

Estimators. Training weights are uniform over observed pairs in Arena48 and category–pair cells in Image. We implement equation 3 through $\Theta = C\Omega_k C^\top$ with factor penalty $\lambda\|C\|_F^2/2$ and bound $\|\Theta\|_{\max} \le 4$, including unobserved entries. The matched BT estimator uses $\Theta_{ij} = r_i - r_j$, centered ratings, the same loss and bound, and penalty $\lambda\sqrt{M}\|r\|_2$. Fits use the training win rates without pseudocounts.

Regularization selection. For each training pair set T, set $\lambda = c\lambda_0$, where $\lambda_0 = \|G_0\|_{\mathrm{op}}$ is recomputed from the training data. The matrix G_0 has entries $(1/2 - \widehat{p}_{ij})/|T|$ on training pairs above the diagonal, opposite entries below, and zeros elsewhere. We select $c \in \{.01, .03, .1, .3, 1\}$ by inner-validation NLL, separately for each estimator and each Image outer fold; differences within 10^{-10} favor larger c. Arena48 selects $c = .03$ for BT/DISC-1 and .1 for DISC-2/4/8. Every Image outer fold selects $c = .01$ for BT/DISC-1 and .03 for DISC-2/4.

Evaluation. Held-out NLL is the unpenalized Bernoulli cross entropy in nats under the same pair or category–pair weights, averaged over the five outer folds for Image. The cyclic fraction is $\|\widehat{\Theta}_{\mathrm{cyc}}\|_F^2/\|\widehat{\Theta}\|_F^2$, reported as a percentage from a separate full-data fit at the selected c. It measures the fitted logit structure outside the additive component. Table 5 reports Image.

Table 6: Arena48 observed-data pipelines before fitted reconstruction. BT/SCO fit observed pairs; QRE/MaxL use Laplace smoothing with unit pseudocounts and zero missing margins. Set recurrence and member retention are distinct descriptive events.

Readout	Point top set	Same set (%)	G in set (%)	CI-separated pairs / 1,128	G rank band
BT	$\{G\}$	99.95	99.95	945	$[1, 1]$
SCO	$\{G, P\}$	38.35	96.15	704	$[1, 5]$
QRE	$\{G\}$	99.80	99.80	720	$[1, 1]$
MaxL first tier	$\{G, P\}$	76.50	99.60	—	—

G: Gemini 2.5 Pro; P: Gemini 2.5 Pro-preview-03-25. QRE uses $\tau = .075$. Counts and bands use 95% percentile score intervals within each pipeline; SCO intervals are solver-conditional. MaxL reports an unordered first tier and has no scalar-score intervals.

D.6 Completion and ranking

We use the Arena48 pool and the BT, DISC-1, DISC-2, and DISC-4 estimators described in Appendix D.4. Point matrices are refitted on all observed pairs.

For downstream Arena48 uncertainty, each of 2,000 draws resamples 133,536 raw votes. The selected regularization ratio c and model pool remain fixed; positive-count edges, proportions, λ_0, the penalty $\lambda = c\lambda_0$, and all four fits are recomputed. Full reconstruction converts every fitted logit to a margin via $\tanh(\Theta_{ij}/2)$. Hybrid completion uses that prediction only on currently missing pairs and keeps the same draw's unsmoothed empirical margins on observed pairs. The two constructions therefore have identical missing-pair predictions draw by draw. The baseline Laplace construction uses $(2w_{ij} - n_{ij})/(n_{ij} + 2)$ on observed pairs and zero on missing pairs. Table 3 summarizes score-interval separation for the declared fitting–completion–QRE pipeline; neither the individual intervals nor the resulting rank bands establish coverage for QRE on the unknown full tournament.

Native BT and SCO fit only currently observed pairs, weighted equally. Table 6 combines these observed-pair pipelines with QRE and MaxL applied to the specified Laplace-completed matrix; it is not four readouts of one common completed matrix. SCO's top tie at $\{G, P\}$ consists of two scores at the upper bound 100. The exact set recurs in 38.35% of draws, while 50.50% return $\{G\}$ alone and G remains a top member in 96.15%. The tie is a numerical score report, not direct evidence that the models have equal strength.

The missing G–P comparison illustrates the scope of indirect evidence. P has 1,320 votes against 22 opponents, all also faced by G, but no votes against G. Native BT separates their score intervals: $[.6719, .7488]$ for G versus $[.3661, .6041]$ for P. Laplace-completed QRE also separates them, with margin-score intervals $[.0898, .1151]$ and $[-.0230, .0476]$. At the fixed point QRE weights, 69.39% of the point score gap comes from opponents whom G has faced and P has not, where P's margins are set to zero. This is an algebraic decomposition at the fitted weights, not the effect of refitting after removing or changing those entries. Separation of these readout scores does not establish a direct G–P win advantage.

The completion analysis uses iterative MaxL: each tier is the union of supports of all maximal lotteries on the remaining subgame, found using coordinate LPs, before that tier is removed. The first-tier support tolerance is 10^{-8}, with frozen feasibility checks and strict-CW shortcuts. This rule differs from selecting one lottery for Table 2. The reported first tier is neither a confidence set nor an internal ranking. Table 3 extracts membership and exact recurrence from all 80 archived chunks covering 2000 resamples.

The DISC-1 comparison identifies how observed-edge reconstruction changes the first tier. Full reconstruction excludes G in 151 draws, selecting a different singleton each time: P in 113, Grok 4 in 19, ChatGPT-4o in 12, and Llama Maverick in seven. On those same 151 draws, hybrid completion gives an expanded tier containing G in 120, $\{G\}$ alone in 16, and excludes G in 15. Among the 113 draws with a fitted P advantage over G, it retains G in 99. One archived example has the cycle $P \to G \to \mathrm{o4mini} \to P$, with effective win rates $.5103, .6091, .5536$; full reconstruction changes the last matchup to favor P, yielding P alone. This is a post-hoc mechanism illustration on a saved draw, not an additional experiment.

Table 7: Auxiliary held-out prediction on separately defined Arena pools. Different test pairs prevent a direct across-column accuracy comparison.

Estimator	Full 53, 241 test pairs	Filtered 48, 227 pairs	Complete 28, 76 pairs
BT	.666391	.665473	.669104
DISC-1	.667071	.665958	.669485
DISC-2	.668335	.667308	.669755
DISC-4	.669515	.667712	.669793

The auxiliary pool analyses show why coverage and pool selection must also be stated. The original 53-model pool has 135,624 votes and 1,203 of 1,378 pairs; the complete 28-model pool has 71,866 votes and all 378 pairs. Their held-out pairs differ, so their raw NLL columns cannot identify an effect of filtering. Table 7 retains these separate evaluations. Restricting both the original full-pool prediction and the Arena48 refit to the same 227 retained test pairs gives BT losses .665491 and .665473, and DISC-1 losses .665975 and .665958, respectively. These are descriptive changes without a significance claim.

The full pool includes a model represented by only two tie votes. Its pure-DISC row is zero, and its point mass is itself a maximal lottery, so it enters the first-tier support union. This describes lack of information under that estimator, not competitive superiority. Filtering changes the pool, observed comparisons, and regularization scale together. On the complete 28-model pool, hybrid completion equals the empirical matrix because no pairs are missing; it provides no separate completion improvement. Across all pools, the bootstrap holds selection choices fixed and cannot remove bias from the unknown missing-edge mechanism.