CONVERGENT ELO: ONLINE PAIRWISE EVALUATION AS ESTIMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Pairwise comparisons are increasingly relied upon to benchmark machine learning systems. When aggregated into a score for a fixed population, these comparisons pose an estimation problem: the reported ratings should settle as evidence accumulates. Fixed-gain Elo acts instead as a tracking algorithm: under noisy stationary comparisons, its ratings continue to fluctuate. We introduce Convergent Elo (CElo), which preserves Elo's two-participant online updates whilst decaying each participant's gain with respect to its own comparison count. We prove almost-sure convergence of the full rating vector under stationary outcome means and any fixed i.i.d. matchmaking distribution with connected support, without requiring global recentring. Further, we identify the limiting relative ratings: under Bradley–Terry realisability, CElo recovers latent strengths up to translation; under misspecification, it converges to the matchmaking-weighted KL-optimal Bradley–Terry approximation. Our analysis establishes a general convergence result for heterogeneous-gain stochastic approximation, deriving stability from coercivity on the identifiable subspace without assuming bounded iterates or imposing projection. In simulations of a stationary population calibrated from Arena Text data, CElo continues to reduce population-target error after fixed-gain Elo plateaus. A controlled ablation shows a clear finite-horizon advantage for participant-local over global gains under heterogeneous participation. Across four chronologically ordered applied datasets, CElo achieves lower prequential BCE than fixed-gain Elo on three, including text and vision model arenas, with no clear difference on the fourth.

1 INTRODUCTION

1.1 MOTIVATION AND PAPER OUTLINE

As machine learning systems are applied to increasingly open-ended tasks, the importance of pairwise comparisons as an evaluation method rises. Take, for instance, LLM code generation. Certain aspects, such as correctness, can be checked systematically, but qualities such as style or maintainability are significantly harder to distil into a single loss. On the other hand, asking a judge (algorithmic or otherwise) to compare two solutions side-by-side and express a relative preference is considerably more tractable. Pairwise evaluation trades *high-bandwidth* standalone assessments of quality for easier-to-obtain *low-bandwidth* preference judgements: a preference for A over B expresses nothing directly about why A is better, or by how much. Therefore, recovering a meaningful population-level score requires many such comparisons, making the choice of aggregation rule highly consequential.

The Elo system (Elo, 1978) is often seen as the canonical aggregation method, valued in part for its simplicity. The central issue is that the fixed gain makes Elo a *tracking* algorithm: it is constantly responsive to new observations. This can be helpful in domains where relative strength fluctuates over time, but in stationary settings with nonzero outcome noise the fixed gain implies a persistent algorithmic performance floor. This is problematic as the output of benchmark aggregators is a score, and that score should be viewed as an *estimator*: it is part of the benchmark output, not merely internal algorithmic state, and crucially, continued evidence should increasingly settle it.

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.

CElo addresses this shortcoming by only adjusting the gain architecture. Each participant tracks their own comparison count, and their gain diminishes with the evidence accumulated about that participant. It is natural to ask what the value of the locality is, as opposed to a diminishing *global* gain based on the total number of comparisons. The principal problem with a global counter is that it decays a participant's gain even on comparisons in which that participant did not appear. A global clock also introduces shared state into an otherwise local two-participant update, requiring synchronisation that CElo (and Elo) avoids.

Proving almost-sure convergence of CElo requires a nonstandard stochastic-approximation argument. Indeed, the participant-local clocks also create the main mathematical difficulty. Coordinates evolve with heterogeneous random gains, which has a warping effect on the trajectory geometry. Further, the Bradley–Terry (BT) objective is translation-invariant, resulting in a noncompact affine minimiser set. Nevertheless, we prove almost-sure convergence of the identifiable component, $\Pi R_n \to z_q^*$; the key insight is to quotient out the translation invariance: coercivity within this identifiable subspace is sufficient to establish stability. A separate lifting argument gives convergence of the unrecentred raw vector, enabling practical implementations of CElo to avoid synchronised recentring. The identifiable result is central to the estimator interpretation, and the raw-state convergence is an architectural strengthening. We also identify the limiting estimand: under BT realisability, matchmaking frequencies do not impact the limiting relative ratings; under misspecification, z_q^* is the matchmaking-weighted KL-optimal BT approximation, thus the comparison policy q is part of the population estimand. These theoretical guarantees apply for a fixed finite population with stationary outcome means and fixed i.i.d. matchmaking with connected support.

CElo is also investigated experimentally. In an Arena-derived stationary population matching the assumptions required for convergence, CElo continues to settle while fixed-gain Elo reaches a persistent error floor. We then isolate the role of participant-local clocks, finding little difference from a global clock when participation is homogeneous but a substantial finite-horizon gap when participation is heterogeneous. Finally, we evaluate CElo on various practical datasets where our stationary assumptions need not hold and find that convergence does not come at an obvious predictive cost.

1.2 Bradley–Terry and Elo

The BT (Bradley & Terry, 1952) model is a standard probabilistic model for pairwise comparisons. Consider a collection of N participants, with participant i assigned a scalar strength parameter θ_i. The probability that i is preferred to j is

$$\mathbb{P}[i \succ j] = \frac{1}{1 + \exp\left(-(\theta_i - \theta_j)\right)}. \tag{1}$$

Only differences between strengths are identifiable: adding the same constant to every θ_i leaves all pairwise probabilities unchanged.

The BT model is most commonly treated as an offline statistical model, with its parameters estimated from a collection of observed comparisons. Elo is closely related to the BT model: seen through a stochastic gradient ascent lens, Elo can be understood as an online BT algorithm. Let r_i and r_j denote the current ratings of two participants, and let $y_i \in \{0, 1\}$ denote the observed outcome from the perspective of i. Using the Bernoulli log-likelihood implied by (1), the stochastic-gradient update for i takes the form

$$r_i^{\text{new}} = r_i^{\text{old}} + K\left(y_i - \frac{1}{1 + \exp\left(-(r_i^{\text{old}} - r_j^{\text{old}})\right)}\right), \tag{2}$$

with the corresponding opposite-signed innovation applied to j. This is the classical Elo update with gain, or K-factor, K.

As discussed above, CElo replaces the fixed K-factor with diminishing participant-local gains, formally defined in Section 2. Unequal local gains need not preserve the global rating offset, even though the BT objective is translation-invariant. Consequently, convergence of rating differences does not by itself imply convergence of the full rating vector: it still leaves open the possibility of unbounded drift along the invariant dimension.

1.3 Related Work

Heterogeneous-gain stochastic approximation. Matrix-step stochastic approximation (SA) provides relevant background for the general convergence theorem of this work. Monnez (2006) addresses random matrix gains and non-singleton solution sets; Bennar & Monnez (2007) consider projected SA on convex sets. Sunehag et al. (2009) provide variable-metric convergence for a unique minimiser under uniform spectral bounds; however, these assumptions do not directly cover CElo's affine minimiser set and participant-local gains (see Section 3.3). Patel et al. (2022) obtain a convergence-or-divergence alternative under spectral Robbins–Monro conditions in a broader nonconvex setting. Theorem 2 instead establishes stability and almost-sure convergence on the identifiable quotient under spectral summability, without assumed iterate boundedness, compact projection, or uniform conditioning.

Participant-local clocks also connect CElo to asynchronous SA. Tsitsiklis (1994) allows asynchronous componentwise updates, but its convergence results are centred on a unique fixed point, with stability assumed or obtained from a stronger contraction-type structure. Relative-frequency ODE methods commonly assume stability (Borkar, 1998; Abounadi et al., 2002); Borkar–Meyn-type results instead derive it using an ODE-at-infinity condition (Borkar & Meyn, 2000; Bhatnagar, 2011; Yu et al., 2026). This criterion is inconclusive for CElo: its drift $h = -\nabla\mathcal{L}_q$ is bounded, so $h(cr)/c \to 0$, and its equilibrium set $z_q^* + \mathrm{span}\{\mathbf{1}\}$ is noncompact. Our analysis establishes quotient stability directly, then separately proves convergence of the unrecentred vector.

Pairwise rating and rank aggregation. For fixed-gain Elo, Aldous (2017) establishes convergence in distribution to stationarity under fixed strengths and uniform random pairing; Oleskey-Taylor & Zanetti (2024) obtain guarantees for time-averaged, projected Elo under a realisable BT model; and Cortez & Tossounian (2026) establish convergence to a stationary distribution. Under misspecification, Tang et al. (2025) study matchmaking-sensitive BT targets and Hamilton et al. (2025) characterise intransitive mean Elo equilibria. Li & Li (2025) develop conditional and marginal KL-based BT strength estimands under covariate shift, including a marginal projection whose covariate-free score equations coincide with those defining our population target. Jarman et al. (2024) prove almost-sure convergence for online rank aggregation, but target rankings induced by comparison inequalities rather than calibrated BT parameters. Glicko and TrueSkill are prominent uncertainty-aware rating systems (Glickman, 1999; 2001; Herbrich et al., 2006); they address related rating problems through different modelling choices rather than the diminishing participant-local gains studied here. Whereas prior work characterises population targets or analyses different procedures, we prove that unprojected participant-local diminishing-gain Elo converges almost surely to the matchmaking-weighted BT target under stationary outcomes and any fixed i.i.d. matchmaking distribution with connected support, including under misspecification.

1.4 Contributions

CElo as online pairwise estimation. Participant-local diminishing gains turn stationary Elo into an almost-surely convergent estimator under fixed i.i.d. matchmaking with connected support. Its limiting target is the centred latent strength vector under BT realisability, and the q-weighted KL-optimal BT approximation under misspecification.

Convergence theory for local heterogeneous gains. Theorem 2 establishes almost-sure identifiable convergence under heterogeneous diagonal gains, deriving stability from coercivity on the identifiable quotient rather than assuming bounded iterates or imposing artificial compact projection. Unlike existing variable-metric convergence results, it does not require a general uniform condition-number bound. A separate CElo-specific argument establishes convergence of the unrecentred rating vector without synchronised recentring.

Finite-horizon and applied evidence. An Arena-derived stationary experiment demonstrates continued settling relative to the fixed-K performance floor, while a paired local/global experiment isolates the finite-horizon benefit of participant-local clocks under heterogeneous participation. On temporally ordered datasets outside the stationary theorem, diminishing participant-local gains do not appear to sacrifice predictive performance.

2 Model and Assumptions

Consider a fixed population of $N \geq 2$ participants, with ratings $R_n \in \mathbb{R}^N$ immediately before comparison $n+1$. Let $G = (V, E)$ be the undirected comparison graph, where $V = \{1, \ldots, N\}$ represents the participants. Fix an arbitrary orientation $e = (i, j)$ for each edge and write $m_e := \hat{e}_i - \hat{e}_j$, where $\hat{e}_i$ is the ith standard basis vector. At comparison $n+1$, an edge E_{n+1} is selected and a score Y_{n+1} is observed from the perspective of its chosen orientation. We write $\mathcal{F}_n$ for the history up to comparison n, including the initial ratings, previous pairings, and their outcomes. The precise filtration and event ordering are given in Appendix A.1.

Let $F(t) = 1/(1+e^{-t})$ denote the logistic CDF. With notation similar to Wong & Shinki (2026), the Elo innovation at match $n+1$ can be written as

$$U_{n+1} := \left(Y_{n+1} - F(m_{E_{n+1}}^\top R_n)\right) m_{E_{n+1}}. \tag{3}$$

Let $\nu_i(n) := \sum_{k=1}^{n} \mathbf{1}\{i \in E_k\}$ count player i's appearances among the first n matches. This is the participant's local clock which is used to compute the gain:

$$\lambda_{n,i} := \frac{c_i}{(a_i + \nu_i(n))^\alpha}, \qquad a_i, c_i > 0, \qquad \frac{1}{2} < \alpha \leq 1. \tag{4}$$

Writing $\Lambda_n := \operatorname{diag}(\lambda_{n,1}, \ldots, \lambda_{n,N})$, the CElo recursion is

$$R_{n+1} = R_n + \Lambda_n U_{n+1}. \tag{5}$$

It should be noted that only ever two ratings are updated after a comparison, and generally by different gains.

Model assumptions. We assume that G is finite and connected. Matchmaking is a fixed distribution: $q = (q_e)_{e \in E}$, where $q_e > 0$ and $\sum_{e \in E} q_e = 1$. At each match, $\mathbb{P}(E_{n+1} = e \mid \mathcal{F}_n) = q_e$ almost surely: pairings may be arbitrarily nonuniform, but matchmaking does not adapt to the ratings or past outcomes. We always have $Y_{n+1} \in [0, 1]$ and, for fixed edge means $p_e \in (0, 1)$,

$$\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n, E_{n+1}] = p_{E_{n+1}} \qquad \text{a.s.} \tag{6}$$

No independence assumption on outcomes is needed beyond these conditional means. The gains follow (4), and R_0 is deterministic and finite. We do not assume BT realisability: the edge means $(p_e)_{e \in E}$ need not arise from a latent strength vector. Appendix A.1 gives the formal specification, including the weaker initial integrability condition sufficient for the proofs.

Population loss. A meaningful loss can be obtained directly from Elo's mean innovation. On edge e, write $\xi = m_e^\top r$. The expected scalar Elo residual is $p_e - F(\xi)$. Naturally, a loss whose *negative* derivative agrees with it must satisfy $\ell_e'(\xi) = F(\xi) - p_e$. Integrating gives

$$\ell_e(\xi) = \log(1 + e^\xi) - p_e \xi = -p_e \log F(\xi) - (1 - p_e) \log\big(1 - F(\xi)\big).$$

Rather than a surrogate chosen to ease analysis, the potential supplied by the expected Elo update is, agreeably, exactly Bernoulli cross-entropy. This recovers, in reverse, the BT-to-Elo connection described in Section 1.2. Averaging over the matchmaking distribution gives

$$\mathcal{L}_q(r) := \sum_{e \in E} q_e \left[-p_e \log F(m_e^\top r) - (1 - p_e) \log\big(1 - F(m_e^\top r)\big)\right]. \tag{7}$$

Under the population comparison process, $\mathcal{L}_q$ is the expected per-match binary log-loss. By the conditional matchmaking and outcome assumptions,

$$\mathbb{E}[U_{n+1} \mid \mathcal{F}_n] = -\nabla \mathcal{L}_q(R_n) \qquad \text{a.s.} \tag{8}$$

Since Λ_n is deterministic given the history, the conditional mean CElo increment is $-\Lambda_n \nabla \mathcal{L}_q(R_n)$: CElo changes the gain architecture governing how this objective is traversed, not the population objective itself. Appendix B records the full derivation of (8) and equivalent forms of $\mathcal{L}_q$.

Finally, since $m_e^\top \mathbf{1} = 0$ for every edge, $\mathcal{L}_q(r + t\mathbf{1}) = \mathcal{L}_q(r)$ for every $r \in \mathbb{R}^N$ and $t \in \mathbb{R}$. This translation symmetry is the source of the non-identifiability addressed in the next section.

3 CONVERGENCE THEORY

Translation invariance poses one of the key challenges in obtaining convergence. If r^* minimises the CElo loss, then so does $r^* + c\mathbf{1}$ for all $c \in \mathbb{R}$. As such, the full population loss cannot have a single minimiser, whereas many standard stochastic-approximation arguments are most naturally stated around a unique equilibrium. Fortunately, the nature of the translation invariance and CElo loss is such that the minima form an affine line in $\mathbb{R}^N$. Consider the centring projection operator $\Pi := I - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$, which conveniently maps onto $\mathcal{H} := \mathbf{1}^\perp = \{r \in \mathbb{R}^N : \mathbf{1}^\top r = 0\}$. Geometrically, this can be thought of as fixing a particular hyperplane perpendicular to the line of equilibria: restricted to this hyperplane, there lies a *single* minimising point. Formally, by Lemma 5, the population loss $\mathcal{L}_q$ has a unique minimiser on $\mathcal{H}$; denote it by $z_q^* := \arg\min_{z \in \mathcal{H}} \mathcal{L}_q(z)$.

Indeed, the overarching approach of this section is to treat the components in $\mathcal{H}$ and $\mathrm{span}\{\mathbf{1}\}$ separately. Considering the projected iterates in $\mathcal{H}$ allows for convergence analysis of *rating differences* (which are identifiable and of central concern). Considering just the $\mathrm{span}\{\mathbf{1}\}$ component allows for convergence analysis of the *rating offset*. Taken together, these imply convergence of the raw rating vector. We stress that the projections here only aid in analysis: CElo itself remains unprojected.

3.1 ALMOST-SURE CONVERGENCE OF CELO

Theorem 1 (Almost-sure convergence of CElo)**.** *Under the model and assumptions of Section 2, there exists a finite real-valued random variable C_∞ such that*

$$R_n \longrightarrow z_q^* + C_\infty \mathbf{1} \qquad \textit{almost surely}. \tag{9}$$

Projecting (9) gives

$$\Pi R_n \longrightarrow z_q^* \qquad \text{almost surely}, \tag{10}$$

so every rating difference and implied BT probability settles, as does the ranking apart from exact ties. This is the conclusion needed to treat the stationary benchmark output as a strongly convergent estimator. The full-state conclusion is stronger, however: it also shows that the common rating offset converges to a finite random limit C_∞. This is non-obvious because, although $U_{n+1} \in \mathbf{1}^\perp$, it is generally the case that $\Lambda_n U_{n+1} \notin \mathbf{1}^\perp$. Nevertheless, raw rating convergence does not require global recentring or compact projection, preserving CElo's participant-local updates.

3.2 STATISTICAL TARGET UNDER MISSPECIFICATION

The key insight from Theorem 1 is that CElo converges to z_q^* up to translation; it is natural then to ask what z_q^* represents. Since no BT realisability assumption was imposed in Section 2, the target must be characterisable even when the true edge means $(p_e)_{e \in E}$ are incompatible with every scalar BT representation.

For $p \in (0, 1)$, let

$$H_{\mathrm{B}}(p) := -p \log p - (1-p)\log(1-p), \qquad C_{p,q} := \sum_{e \in E} q_e H_{\mathrm{B}}(p_e).$$

Proposition 1 (Pseudo-true Bradley–Terry target)**.** *The population loss admits the decomposition*

$$\mathcal{L}_q(r) = C_{p,q} + \sum_{e \in E} q_e D_{\mathrm{KL}}\big(\mathrm{Bern}(p_e) \,\big\|\, \mathrm{Bern}(F(m_e^\top r))\big)\,. \tag{11}$$

Consequently, z_q^* is the unique centred Bradley–Terry parameter minimising the q-weighted Bernoulli KL discrepancy. In particular:

(i) *if there exists $\theta \in \mathbb{R}^N$ such that*

$$p_e = F(m_e^\top \theta) \qquad \forall e \in E,$$

then

$$z_q^* = \Pi\theta \tag{12}$$

for every strictly positive matchmaking distribution on the same connected support graph;

(ii) *without Bradley–Terry realisability, the minimiser z_q^* can depend on q.*

Under realisability, all KL terms can vanish simultaneously, so the matchmaking distribution q impacts sampling efficiency but not the target; Theorem 1 yields $\Pi R_n \to \Pi\theta$ and $R_n \to \Pi\theta + C_\infty \mathbf{1}$ almost surely. Under misspecification, however, the constraints can conflict. In this case, the weight of a given comparison edge q_e determines its relative importance in the best scalar BT approximation. Therefore, when the scalar BT model is misspecified (for instance, under forms of intransitivity that are incompatible with scalar BT strengths), changing the matchmaking policy can change the population quantity being summarised. Appendix C gives the proof and an explicit three-player example, and Appendix E.3 gives a controlled finite-run illustration.

3.3 Convergence on the identifiable quotient

The CElo update (5) is a particular stochastic-approximation recursion with a diagonal gain matrix. To isolate the part of the convergence argument that does not depend on pairwise comparisons or participant-local clocks, consider instead

$$X_{n+1} = X_n + \Lambda_n H_{n+1} \tag{13}$$

in $\mathbb{R}^d$. Here X_n is adapted to a history $\mathcal{F}_n$, H_{n+1} is the next conditionally square-integrable innovation, and $\Lambda_n = \mathrm{diag}(\lambda_{n,1}, \dots, \lambda_{n,d})$ is an $\mathcal{F}_n$-measurable diagonal matrix with strictly positive entries. Write $\underline{\lambda}_n := \min_i \lambda_{n,i}$ and $\overline{\lambda}_n := \max_i \lambda_{n,i}$. This extends the usual scalar Robbins–Monro stepsize (Robbins & Monro, 1951) to coordinate-dependent gains; Appendix A.2 specifies the general filtered setup. CElo is recovered by taking $d = N$, $X_n = R_n$, and $H_{n+1} = U_{n+1}$, but the following result permits more general smooth objectives and linear symmetries under the conditions below.

Theorem 2 (Almost-sure convergence modulo a linear symmetry)**.** *Let P be an orthogonal projection on $\mathbb{R}^d$ and let $\mathcal{H}_P := \mathrm{range}(P)$. Suppose $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable with L-Lipschitz gradient and satisfies*

$$\mathcal{L}(x) = \mathcal{L}(Px) \qquad \forall x \in \mathbb{R}^d. \tag{14}$$

Assume that $\mathcal{L}|_{\mathcal{H}_P}$ is coercive and strictly convex, and write

$$z^* := \operatorname*{arg\,min}_{z \in \mathcal{H}_P} \mathcal{L}(z).$$

For the recursion (13), assume

$$\mathbb{E}[H_{n+1} \mid \mathcal{F}_n] = -\nabla\mathcal{L}(X_n) \qquad \text{a.s.}, \tag{15}$$

and, almost surely,

$$\sum_{n=0}^{\infty} \underline{\lambda}_n = \infty, \qquad \sum_{n=0}^{\infty} \overline{\lambda}_n^2 < \infty. \tag{16}$$

Suppose also that $\mathbb{E}[\mathcal{L}(X_0) - \mathcal{L}(z^*)] < \infty$ and that, for deterministic constants $A, B \geq 0$,

$$\mathbb{E}\Big[\|H_{n+1}\|^2 \mid \mathcal{F}_n\Big] \leq A + B\big(\mathcal{L}(X_n) - \mathcal{L}(z^*)\big) \qquad \text{a.s.} \tag{17}$$

Then

$$PX_n \longrightarrow z^* \qquad \text{almost surely}. \tag{18}$$

Equivalently, the distance from X_n to the affine minimiser set $z^* + \ker(P)$ converges to zero almost surely.

Proof idea. We build upon the argument found in Duflo (1997, Theorem 1.4.26). Broadly, we aim to establish that the Lyapunov function $V_n := \mathcal{L}(PX_n) - \mathcal{L}(z^*)$ converges specifically to zero. The descent lemma and diagonal-gain bounds yield a Robbins–Siegmund-style inequality, implying that V_n converges, though not yet necessarily to zero, and that an appropriately weighted gradient norm is summable. If the limit of V_n were positive, quotient coercivity would confine PX_n to a compact sublevel annulus on which the gradient is bounded away from zero. This contradicts the divergence of the effective descent weights, hence the limit must be zero. Appendix C gives the full proof.

The closest result for our purposes is that of Sunehag et al. (2009). Quotienting removes the non-uniqueness of the CElo objective, but does not make their theorem directly applicable: their matrix step is a scalar Robbins–Monro gain multiplied by a metric whose eigenvalues are uniformly bounded above and away from zero. Participant-local clocks do not provide such uniform conditioning; the relative gains can be arbitrarily imbalanced along finite sample paths, even though they become comparable almost surely asymptotically. Theorem 2 instead works directly with the eigenvalue sums in (16), deriving stability from quotient coercivity without a uniform condition-number assumption.

For CElo, quotient coercivity has a direct graph-theoretic origin. Let $r \in \mathcal{H}$ be an arbitrary centred rating vector. The proof strategy relies on exploiting the initially unassuming fact that $\max_i r_i \geq 0 \geq \min_i r_i$. From here, we consider a path between a maximally and minimally rated participant, which is ensured by connectivity, and derive a lower bound for $\mathcal{L}_q(r)$ in terms of $\max_i r_i - \min_i r_i$, and then an upper bound for $\|r\|$ in terms of the same difference. Chaining the inequalities implies coercivity. The full argument is Lemma 5.

We can now specialise the theorem to CElo. Set $P = \Pi$, $X_n = R_n$, $H_{n+1} = U_{n+1}$, and $\mathcal{L} = \mathcal{L}_q$. Equation (8), the bound $\|U_{n+1}\|^2 \leq 2$, Lemma 5, and Lemma 6 verify the assumptions for $1/2 < \alpha \leq 1$; in particular, the second-moment bound (17) holds with $A = 2$ and $B = 0$. Therefore we obtain:

Corollary 1 (Centred convergence)**.** *Under the model of Section 2,*

$$\Pi R_n \longrightarrow z_q^* \qquad \textit{almost surely}. \tag{19}$$

3.4 Unrecentred point convergence

Corollary 1 addresses subspace $\mathcal{H}$, but says nothing about $\mathrm{span}\{\mathbf{1}\}$. Analysis in this direction is challenging as player-local gains do not ensure equal-and-opposite changes in rating, which fixed-gain Elo does. This introduces a drift within $\mathrm{span}\{\mathbf{1}\}$ which has no restoring gradient to prevent unbounded wandering. Appendix C.2 shows that the key is to weigh the gains in terms of participation frequency such that

$$\Lambda_n w = n^{-\alpha}\mathbf{1} + r_n, \qquad n \geq 1, \quad \sum_{n=1}^{\infty} \|r_n\| < \infty.$$

The leading term cancels in the update of $w^\top R_n$, while bounded innovations and the summable remainder ensure convergence, yielding a finite $C_\infty \mathbf{1}$ target. This is again achieved without synchronised recentring, or violating locality.

4 Experiments

We aim to experimentally validate the behaviour of CElo in three different key settings. The first illustrates CElo's finite-horizon settling behaviour relative to Elo on a stationary population calibrated from Arena Text data. The second is an ablation study comparing participant-local clocks against a global clock, characterising performance in highly homogeneous and heterogeneous scenarios showing when participant-local clocks offer an advantage. Finally, the predictive performance of CElo is tested against common rating systems on four practical evaluation datasets including Text and Vision Arenas.

4.1 Finite-horizon settling on an Arena-derived stationary population

Does CElo continue to settle after fixed-gain Elo reaches its error floor, and how does the choice of gain-decay exponent α affect this behaviour? To investigate this, we construct a stationary population from Arena Text data. Our matchmaking distribution q matches that of the dataset empirically, and a smoothed edge mean defines p_e. We numerically compute the corresponding centred target z_q^*. The frozen empirical edge means are not exactly BT-realisable (Figure 3), hence z_q^* is the q-weighted pseudo-true target of Proposition 1.

We generate 32 independent i.i.d. comparison histories from this fixed population. We compare CElo across $\alpha \in \{0.51, 0.60, 0.75, 0.90, 1.00\}$ against validation-selected fixed-K Elo, using common initial calibration. Full experimental details are given in Appendix E.1.

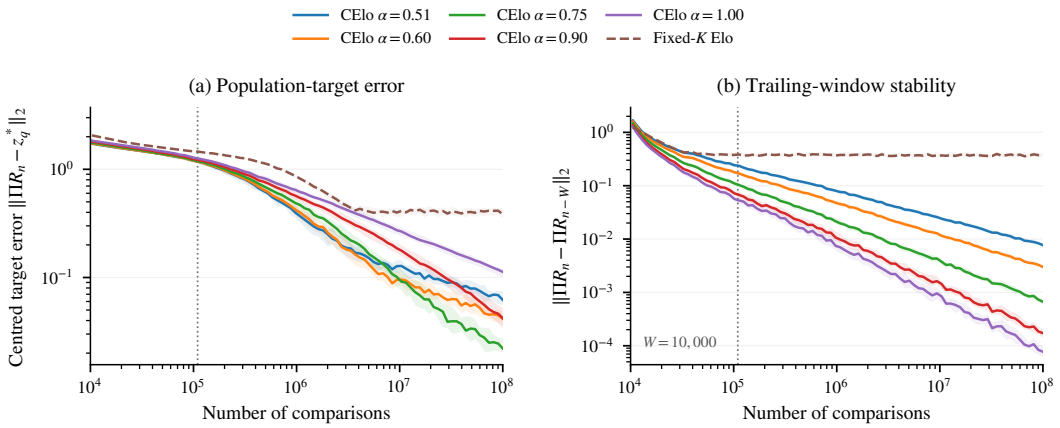

Figure 1: Finite-horizon settling on the Arena-derived stationary population over 32 i.i.d. resampled streams. Curves show medians with 20–80% ranges. Panel (a) shows centred target error $\|\Pi R_n - z_q^*\|_2$; panel (b) shows trailing-window movement $\|\Pi R_n - \Pi R_{n-W}\|_2$, with $W = 10{,}000$. The dotted line marks the 109,646 original Arena comparisons used to calibrate the population. All plotted curves use simulated draws from this frozen population.

Figure 1 shows fixed-K Elo reaching a plateau in both target error and trailing-window movement, while every CElo schedule continues to settle over the displayed horizon. Larger α suppresses movement faster, but interestingly does not uniformly improve target error: $\alpha = 0.75$ achieves the lowest late target error here, while $\alpha = 1$ yields the greatest stability. This illustrates a finite-horizon trade-off between estimation accuracy and stability, rather than a universally preferable decay schedule.

4.2 Participant-local versus global diminishing gains ablation

When are participant-local clocks advantageous over a globally diminishing gain? The concern with a global clock is that a participant's gain diminishes even when they are not involved in a comparison. We investigate the effect of this distinction in a stationary BT-realisable population, comparing CElo against a global clock baseline. For comparison $n + 1$, the respective gains are

$$\lambda_{n,i}^{\text{local}} = \frac{c}{(M + \nu_i(n))^\alpha}, \qquad \lambda_n^{\text{global}} = \frac{c}{(M + 2n/N)^\alpha}.$$

The global clock is scaled by $2/N$ to match the expected local participation count under homogeneous exposure. Both methods use identical initialisation and calibration ($\alpha = 0.75$), and receive the same comparisons and outcomes.

Figure 2 shows that the two methods behave almost identically under homogeneous participation, with median errors of approximately 0.035 after 10^7 comparisons. Under strongly heterogeneous participation, however, the difference is unmistakable: CElo achieves an error of 0.063, compared with the global clock's 0.950. This demonstrates the finite-horizon advantage of mechanically relating each participant's gain to their own accumulated evidence.

4.3 Applied predictive evaluation

Does CElo exhibit a predictive cost on real comparison datasets, where stationarity need not hold? We compare CElo prequentially against fixed-gain Elo on four datasets: Arena Text, Arena Vision, BASIL (StarCraft agent matches), and TCEC (chess engine games). Full dataset descriptions are given in Appendix E.5. Unlike the stationary experiments above, we retain the chronological ordering of the datasets. These experiments therefore test predictive performance outside the assumptions of our convergence theorem.

CElo obtains lower observed BCE on Arena Text, Arena Vision, and BASIL. The corresponding paired moving-block-bootstrap 95% intervals for the difference lie entirely below zero, while the

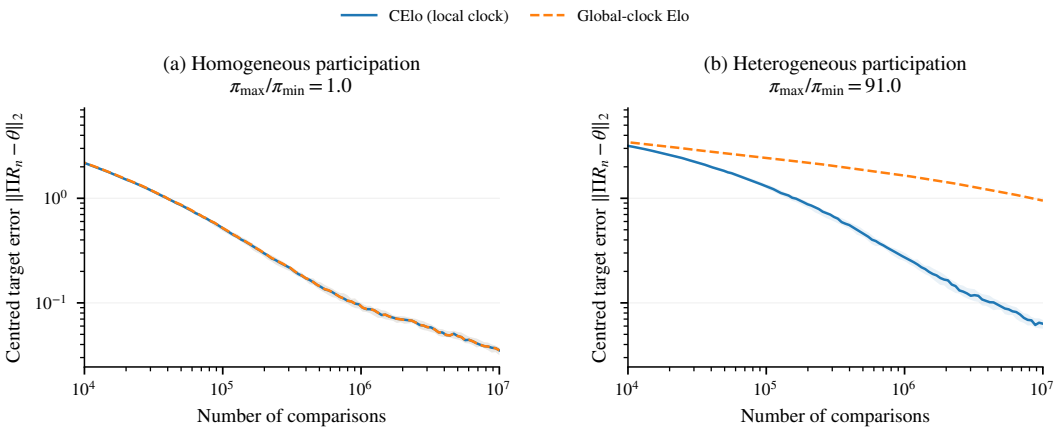

Figure 2: Participant-local versus globally clocked diminishing gains in a stationary BT-realisable population. Curves show medians over 32 paired streams, with 20–80% ranges that are largely obscured by the lines. Panel (a) has homogeneous participation ($\pi_{\max}/\pi_{\min} = 1.0$), while panel (b) has strongly heterogeneous participation ($\pi_{\max}/\pi_{\min} = 91.0$). The metric is centred target error $\|\Pi R_n - \theta\|_2$.

interval for TCEC contains zero. We interpret these intervals descriptively, given the potentially nonstationary evaluation streams. CElo also achieves lower Brier scores (Brier, 1950) on all four datasets. Within these experiments, diminishing gains therefore do not appear to impose a consistent predictive cost, and frequently improve observed predictive performance.

Notably, on BASIL, validation selects $\alpha = 0.50023$. This is explainable by the $1/2$ boundary resulting in larger gains allowing for more adaptation. Exact results and broader method comparisons are provided in Appendix F, including exploratory CElo variants that further improve BCE through participant-specific gain resets.

5 Conclusion and Future Work

The Elo rating system was developed over 60 years ago as a heuristic for determining the relative strength of chess players. Viewed retrospectively through the framework of this work, it is evident that Elo was an early distributed online learning algorithm. Whilst it has stood the test of time in tracking applications, fixed-population ML benchmarks demand a different requirement: the reported score should increasingly settle as evidence accumulates.

CElo achieves this through changing the gain architecture alone. By exploiting coercivity on the identifiable quotient, we establish stability and almost-sure convergence of relative ratings under heterogeneous participant-local gains. A separate argument establishes convergence of the unrecentred raw rating vector. Our experiments demonstrate continued settling on an Arena-derived stationary population; the finite-horizon advantage of participant-local clocks under heterogeneous participation; and CElo's predictive viability on practical datasets where stationarity need not hold.

Finite-time theory, adaptive matchmaking, and rigorous nonstationary extensions of CElo present avenues for future work. Moreover, the heterogeneous-gain theory developed in this work may also be applicable to uncertainty-aware rating systems in regimes where their gains diminish. More broadly, deterministic local clocks, Bayesian uncertainty, and change-detection resets are all mechanisms for controlling participant-specific adaptation; understanding when they result in convergent or tracking behaviour is a natural extension.

Reproducibility statement

Appendices A–C provide the complete model specification and proofs of the theoretical claims. Appendix D gives algorithmic details. Appendix E documents the experimental protocols, population construction, preprocessing, hyperparameter optimisation, uncertainty estimation, and datasets. Appendix F reports the complete experimental results.

AI use statement

The core theory and conceptualisation of this work were developed without generative AI assistance. Generative AI was subsequently used to assist in generalising certain theoretical results, in particular going from approximately uniform to non-uniform matchmaking policies. It was also used as an aid in checking mathematical arguments and in editing proof write-ups for clarity and presentation. All mathematical claims and proofs were human-verified.

Generative AI was used during the writing process. An original draft of the paper was provided to an LLM and used to produce a more polished draft under iterative human direction. Regardless, the final manuscript was then extensively rewritten by hand with AI providing general feedback, while retaining some of the organisation suggested, and a minority of the phrasing suggested, during this process. AI assistance was also used in editing more formulaic parts of the exposition, such as the abstract.

The core algorithms were implemented by hand without generative AI assistance and were subsequently tested for correctness. Given these reference implementations, generative AI was used substantially to assist with the implementation of the synthetic experiments and with the production of tables and figures. All AI-generated or AI-assisted code and experimental artefacts used in the paper were reviewed and human-verified.

Generative AI was not used to conceptualise the research question, analyse the experimental results, interpret those results, or draw the conclusions reported in the paper. The central literature review was conducted manually; generative AI was additionally used to assist in searching for relevant related work.

We take responsibility for the final content of this work, including all text, claims, code, and artefacts produced with the assistance of generative AI.

References

Jinane Abounadi, Dimitri P. Bertsekas, and Vivek S. Borkar. Stochastic approximation for nonexpansive maps: Application to Q-learning algorithms. *SIAM Journal on Control and Optimization*, 41(1):1–22, 2002. doi: 10.1137/S0363012998346621.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631. ACM, 2019. doi: 10.1145/3292500.3330701.

David Aldous. Elo ratings and the sports model: A neglected topic in applied probability? *Statistical Science*, 32(4):616–629, November 2017. doi: 10.1214/17-STS628.

BASIL Ladder. BASIL Ladder. Website, 2026. URL https://www.basil-ladder.net/. Accessed 2026-09-13.

Abdelkrim Bennar and Jean-Marie Monnez. Almost sure convergence of a stochastic approximation process in a convex set. *International Journal of Applied Mathematics*, 20(5):713–722, 2007.

James S. Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*, volume 24, pp. 2546–2554, 2011.

Shalabh Bhatnagar. The Borkar–Meyn theorem for asynchronous stochastic approximations. *Systems & Control Letters*, 60(7):472–478, 2011. doi: 10.1016/j.sysconle.2011.04.002.

Vivek S. Borkar. Asynchronous stochastic approximations. *SIAM Journal on Control and Optimization*, 36(3):840–851, 1998. doi: 10.1137/S0363012995282784.

Vivek S. Borkar and Sean P. Meyn. The O.D.E. method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469, 2000. doi: 10.1137/S0363012997331639.

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

Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950. doi: 10.1175/1520-0493(1950)078⟨0001:VOFEIT⟩2.0.CO;2.

Roberto Cortez and Hagop Tossounian. Convergence and stationary distribution of Elo rating systems, 2026. URL https://arxiv.org/abs/2410.09180.

Marie Duflo. *Random Iterative Models*, volume 34 of *Applications of Mathematics*. Springer, Berlin, 1997. ISBN 978-3-540-57100-1.

Arpad E. Elo. *The Rating of Chessplayers, Past and Present*. Batsford, London, 1978. ISBN 978-0-7134-1860-6.

Stefan Falkner, Aaron Klein, and Frank Hutter. BOHB: Robust and efficient hyperparameter optimization at scale. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1437–1446. PMLR, 2018.

Mark E. Glickman. Parameter estimation in large dynamic paired comparison experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(3):377–394, 1999. doi: 10.1111/1467-9876.00159.

Mark E. Glickman. Dynamic paired comparison models with stochastic variances. *Journal of Applied Statistics*, 28(6):673–689, 2001. doi: 10.1080/02664760120059219.

Peter Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980. ISBN 978-0-12-319350-6.

Adam H. Hamilton, Anna Kalenkova, and Matthew Roughan. The impact of intransitivity on the Elo rating system. *PLOS ONE*, 20(12):e0338261, 2025. doi: 10.1371/journal.pone.0338261.

Ralf Herbrich, Tom Minka, and Thore Graepel. TrueSkill: A Bayesian skill rating system. In *Advances in Neural Information Processing Systems*, volume 19, pp. 569–576. MIT Press, 2006.

Benjamin Jarman, Lara Kassab, Deanna Needell, and Alexander Sietsema. Stochastic iterative methods for online rank aggregation from pairwise comparisons. *BIT Numerical Mathematics*, 64(3):26, 2024. doi: 10.1007/s10543-024-01024-x.

J. L. Kelly, Jr. A new interpretation of information rate. *The Bell System Technical Journal*, 35(4):917–926, 1956. doi: 10.1002/j.1538-7305.1956.tb03809.x.

Hans R. Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989. doi: 10.1214/aos/1176347265.

Xiudi Li and Sijia Li. Efficient inference for covariate-adjusted Bradley–Terry model with covariate shift, 2025. URL https://arxiv.org/abs/2503.18256.

LMArena AI. Arena Human Preference 140K. Hugging Face dataset, 2025a. URL https://huggingface.co/datasets/lmarena-ai/arena-human-preference-140k. Accessed 2026-09-13.

LMArena AI. VisionArena-Battle. Hugging Face dataset, 2025b. URL https://huggingface.co/datasets/lmarena-ai/VisionArena-Battle. Accessed 2026-09-13.

Jean-Marie Monnez. Almost sure convergence of stochastic gradient processes with matrix step sizes. *Statistics & Probability Letters*, 76(5):531–536, March 2006.

Sam Olesker-Taylor and Luca Zanetti. An analysis of Elo rating systems via Markov chains. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-4389.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954. doi: 10.1093/biomet/41.1-2.100.

Vivak Patel, Shushu Zhang, and Bowen Tian. Global convergence and stability of stochastic gradient descent. In *Advances in Neural Information Processing Systems*, volume 35. Curran Associates, Inc., 2022. doi: 10.52202/068431-2610.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951. doi: 10.1214/aoms/1177729586.

Herbert Robbins and David Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In Jagdish S. Rustagi (ed.), *Optimizing Methods in Statistics*, pp. 233–257. Academic Press, New York, 1971. doi: 10.1016/B978-0-12-604550-5.50015-8.

Peter Sunehag, Jochen Trumpf, S. V. N. Vishwanathan, and Nicol N. Schraudolph. Variable metric stochastic approximation theory. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pp. 560–566. PMLR, 2009.

Shange Tang, Yuanhao Wang, and Chi Jin. Is Elo rating reliable? A study under model misspecification, 2025. URL https://arxiv.org/abs/2502.10985.

TCEC. Top Chess Engine Championship. Website, 2026. URL https://www.tcec-chess.com/. Accessed 2026-09-13.

John N. Tsitsiklis. Asynchronous stochastic approximation and Q-learning. *Machine Learning*, 16 (3):185–202, 1994. ISSN 1573-0565. doi: 10.1023/A:1022689125041. URL https://doi.org/10.1023/A:1022689125041.

Kin Hang Wong and Kazuhiko Shinki. Pairwise-Elo rating system. *Journal of Quantitative Analysis in Sports*, 22(3):263–305, 2026. doi: 10.1515/jqas-2025-0150. URL https://doi.org/10.1515/jqas-2025-0150.

Huizhen Yu, Yi Wan, and Richard S. Sutton. Asynchronous stochastic approximation with applications to average-reward reinforcement learning. *SIAM Journal on Control and Optimization*, 64 (3):1456–1481, 2026. doi: 10.1137/25M1769806.

A Formal Model Specifications

A.1 Formal CElo specification

Section 2 defines the comparison graph, Elo innovation, local-clock gains, and rating recursion. Here we make the conditioning and match timing explicit. For $G = (V, E)$ and the chosen edge orientations of Section 2, let R_n be the rating vector immediately before match $n + 1$ and define

$$\mathcal{F}_n := \sigma(R_0, E_1, Y_1, \dots, E_n, Y_n), \qquad \mathcal{G}_{n+1} := \mathcal{F}_n \vee \sigma(E_{n+1}),$$

where $\sigma(\cdot)$ is taken to mean the sigma-field generated by its arguments. Hence $\mathcal{F}_n$ is the full history after match n, and $\mathcal{G}_{n+1}$ additionally reveals the next selected edge, but not its outcome. For each $n \geq 0$ we assume the following conditions:

A1. $G = (V, E)$ is finite and connected, with $V = \{1, \dots, N\}$ and $N \geq 2$. The orientation $e = (i, j)$ of an edge fixes $m_e = \hat{e}_i - \hat{e}_j$ and the perspective of its score.

A2. Matchmaking has a fixed distribution $q = (q_e)_{e \in E}$ with $q_e > 0$ and $\sum_{e \in E} q_e = 1$. Formally,

$$\mathbb{P}(E_{n+1} = e \mid \mathcal{F}_n) = q_e \qquad \text{a.s.} \tag{20}$$

The edges $(E_n)_{n \geq 1}$ are therefore i.i.d., with no requirement of uniform matchmaking or a complete graph.

A3. The scores $Y_{n+1} \in [0,1]$ have fixed conditional means $p_e \in (0,1)$:

$$\mathbb{E}[Y_{n+1} \mid \mathcal{G}_{n+1}] = p_{E_{n+1}} \qquad \text{a.s.} \tag{21}$$

No temporal independence is assumed for the outcomes beyond this conditional-mean requirement.

A4. The update is (3)–(5), with the local counts and positive gains in (4). In the main text, R_0 is assumed deterministic and finite. However, the proofs require a strictly weaker initial-loss condition

$$\mathbb{E}[\mathcal{L}_q(R_0)] < \infty. \tag{22}$$

When using this weaker condition, R_0 is understood to be finite almost surely and $\mathcal{F}_0$-measurable.

BT realisability is not assumed: the fixed edge means $(p_e)_{e\in E}$ may be incompatible with every latent strength vector. At the start of match $n+1$, R_n, the counts $\nu_i(n)$, and the diagonal gain Λ_n are $\mathcal{F}_n$-measurable. The edge E_{n+1} is then selected, the score Y_{n+1} observed, and the innovation U_{n+1} formed. Applying the pre-match gain Λ_n yields R_{n+1}; only then do the counts advance to $\nu_i(n+1)$. In particular, U_{n+1} is $\mathcal{F}_{n+1}$-measurable and $\|U_{n+1}\|^2 \le 2$.

The empirical schedule $\lambda_{n,i} = c/(M + \nu_i(n))^\alpha$ is the special case $c_i = c$ and $a_i = M$ for every player.

A.2 Generic diagonal-gain recursion

Theorem 2 is stated for an abstract filtered probability space with a filtration $(\mathcal{F}_n)_{n\ge 0}$, not necessarily the comparison history defined in A.1. For an arbitrary dimension d, the recursion (13) has $X_n \in \mathbb{R}^d$ finite and $\mathcal{F}_n$-measurable, and $H_{n+1} \in \mathbb{R}^d$ is $\mathcal{F}_{n+1}$-measurable with

$$\mathbb{E}[\|H_{n+1}\|^2 \mid \mathcal{F}_n] < \infty \qquad \text{a.s.}$$

The gain matrix $\Lambda_n = \operatorname{diag}(\lambda_{n,1}, \dots, \lambda_{n,d})$ is $\mathcal{F}_n$-measurable, with every $\lambda_{n,i} > 0$ almost surely. Its extremal diagonal entries are $\underline{\lambda}_n = \min_i \lambda_{n,i}$ and $\overline{\lambda}_n = \max_i \lambda_{n,i}$. The mean-field, sum, and moment conditions required for convergence are stated in Theorem 2; in particular, the theorem does not assume that the gains arise from participation counts or that the innovation has pairwise support. When applying the theorem to CElo, the filtration is the one in Appendix A.1, $d = N$, $X_n = R_n$, and $H_{n+1} = U_{n+1}$.

B Elo Population Loss

This appendix explores equivalent forms of the CElo population loss and its relation to the mean update. Write

$$\xi_e(r) := m_e^\top r, \qquad s_e(r) := F(\xi_e(r)).$$

The main-text objective in (7) is the matchmaking-weighted binary cross-entropy

$$\mathcal{L}_q(r) = \sum_{e\in E} q_e \left[-p_e \log s_e(r) - (1-p_e) \log\big(1 - s_e(r)\big) \right].$$

Applying a standard result for the logistic link, each edgewise term has the equivalent *softplus* form

$$-p \log F(\xi) - (1-p) \log\big(1 - F(\xi)\big) = \log(1 + e^\xi) - p\xi.$$

Hence

$$\mathcal{L}_q(r) = \sum_{e\in E} q_e \left[\log\big(1 + \exp(m_e^\top r)\big) - p_e m_e^\top r \right]. \tag{23}$$

The cross-entropy form makes the statistical meaning of the objective explicit, while (23) is often convenient for the differentiation and geometric arguments used below.

Mean update as a stochastic gradient. Since

$$\frac{\mathrm{d}}{\mathrm{d}\xi}\left[\log(1+e^{\xi}) - p\xi\right] = F(\xi) - p,$$

differentiating (23) gives

$$\nabla \mathcal{L}_q(r) = \sum_{e\in E} q_e\big(F(m_e^\top r) - p_e\big)m_e. \tag{24}$$

On the other hand, applying fixed matchmaking, (6), and the definition of conditional expectation yields

$$\begin{aligned}\mathbb{E}[U_{n+1} \mid \mathcal{F}_n] &= \sum_{e\in E} q_e\big(p_e - F(m_e^\top R_n)\big)m_e \\ &= -\nabla \mathcal{L}_q(R_n) \qquad \text{a.s.}\end{aligned} \tag{25}$$

Thus CElo is a stochastic-gradient recursion for the population cross-entropy, with participant-local coordinate gains encoded by Λ_n. Equivalently, the cross-entropy is not an externally chosen surrogate: on one edge the mean Elo innovation is $p_e - F(\xi)$, so a scalar potential whose negative derivative matches that mean update must satisfy $\ell_e'(\xi) = F(\xi) - p_e$; integrating gives the same logistic cross-entropy up to an additive constant.

KL form. For any $p, s \in (0,1)$,

$$-p\log s - (1-p)\log(1-s) = H_{\mathrm{B}}(p) + D_{\mathrm{KL}}(\mathrm{Bern}(p) \,\|\, \mathrm{Bern}(s))\,.$$

Applying this identity edgewise and averaging with weights q_e yields the decomposition in Proposition 1. In particular, the entropy contribution depends only on the fixed edgewise means $(p_e)_{e\in E}$, so minimising $\mathcal{L}_q$ is equivalent to minimising the matchmaking-weighted Bernoulli KL discrepancy.

A log-optimal betting interpretation (binary outcomes only). There is also a betting interpretation of the KL discrepancy. Suppose an idealised zero-vig bookmaker quotes reciprocal odds $1/s_e(r)$ and $1/(1-s_e(r))$ for the two outcomes on edge e, although their true probabilities are p_e and $1-p_e$. A bettor who knows p_e and allocates fractions b and $1-b$ of their wealth to the two outcomes maximises expected log-growth at the Kelly allocation $b = p_e$ (Kelly, 1956). The resulting maximum is

$$\max_{0<b<1}\left\{p_e\log\frac{b}{s_e(r)} + (1-p_e)\log\frac{1-b}{1-s_e(r)}\right\} = D_{\mathrm{KL}}(\mathrm{Bern}(p_e) \,\|\, \mathrm{Bern}(s_e(r)))\,.$$

Thus a bookmaker quoting mispriced BT probabilities exposes itself to precisely the edgewise KL discrepancy that Elo's population objective minimises. Averaged over matchmaking, the bettor's maximum expected log-wealth-growth is $\mathcal{L}_q(r) - C_{p,q}$. This is *not* $\mathcal{L}_q(r)$ itself: the full loss includes the fixed entropy term $C_{p,q}$. Of course, despite the functionals differing, their minimising rating vectors coincide. This observation applies to binary outcomes and is provided as an intuitive argument for what CElo is optimising; it is not needed for the formal convergence argument.

C Statements and Proofs

Throughout this appendix, statements involving random variables are understood to hold almost surely unless noted otherwise. Vector norms are Euclidean, and $\|\cdot\|_{\mathrm{op}}$ denotes the induced Euclidean (spectral) matrix norm. The principal results are stated once in the main text; this appendix provides their proofs, as well as the supporting theorems and lemmas needed to make the argument self-contained. The order of the results and proofs in the appendix differs to the presentation in the main text. The appendix ordering is more natural for constructing the arguments from scratch. We begin, for convenience, with the standard Robbins–Siegmund almost-supermartingale lemma (Robbins & Siegmund, 1971) stated without proof.

Lemma 1 (Robbins–Siegmund almost-supermartingale lemma; version used here)**.** *Let (V_n), (U_n), (A_n) and (B_n) be nonnegative adapted random variables satisfying*

$$\mathbb{E}[V_{n+1} \mid \mathcal{F}_n] \le (1+A_n)V_n - U_n + B_n \quad \text{a.s.}$$

If $\sum_n A_n < \infty$ and $\sum_n B_n < \infty$ almost surely, then V_n converges almost surely to a finite random variable and $\sum_n U_n < \infty$ almost surely.

This lemma is a classical result and underpins the core convergence theorem.

C.1 Convergence under coordinate-local gains in the projected subspace

Proof of Theorem 2. Equation (14) and differentiability imply

$$\nabla\mathcal{L}(x) = \nabla\mathcal{L}(Px) \quad \text{and} \quad P\nabla\mathcal{L}(x) = \nabla\mathcal{L}(x). \tag{26}$$

Indeed, $\mathcal{L} = \mathcal{L} \circ P$ gives $\nabla\mathcal{L}(x) = P\nabla\mathcal{L}(Px)$, because $P = P^\top$; evaluating at Px gives $\nabla\mathcal{L}(Px) = P\nabla\mathcal{L}(Px)$, and the claim follows.

Set

$$V_n := \mathcal{L}(X_n) - \mathcal{L}(z^*) = \mathcal{L}(PX_n) - \mathcal{L}(z^*) \geq 0.$$

By the descent lemma and (13),

$$\mathcal{L}(X_{n+1}) \leq \mathcal{L}(X_n) + \langle \nabla\mathcal{L}(X_n), \Lambda_n H_{n+1}\rangle + \frac{L}{2}\left\|\Lambda_n H_{n+1}\right\|^2.$$

Taking conditional expectations and using (15),

$$\begin{aligned}\mathbb{E}[V_{n+1} \mid \mathcal{F}_n] &\leq V_n - \langle \nabla\mathcal{L}(X_n), \Lambda_n \nabla\mathcal{L}(X_n)\rangle + \frac{L}{2}\mathbb{E}[\|\Lambda_n H_{n+1}\|^2 \mid \mathcal{F}_n] \\ &\leq V_n - \underline{\lambda}_n \left\|\nabla\mathcal{L}(X_n)\right\|^2 + \frac{L}{2}\overline{\lambda}_n^2 \mathbb{E}[\|H_{n+1}\|^2 \mid \mathcal{F}_n]. \end{aligned} \tag{27}$$

By (17), (27) becomes

$$\mathbb{E}[V_{n+1} \mid \mathcal{F}_n] \leq \left(1 + \frac{LB}{2}\overline{\lambda}_n^2\right) V_n - \underline{\lambda}_n \left\|\nabla\mathcal{L}(X_n)\right\|^2 + \frac{LA}{2}\overline{\lambda}_n^2.$$

Lemma 1 and (16) imply that $V_n \to V_\infty < \infty$ almost surely and

$$\sum_n \underline{\lambda}_n \left\|\nabla\mathcal{L}(X_n)\right\|^2 < \infty \quad \text{a.s.} \tag{28}$$

It remains to show $V_\infty = 0$. Let Ω_0 be the intersection of the probability-one events on which the preceding Robbins–Siegmund conclusions and the step-size assumptions hold. Then $\mathbb{P}(\Omega_0) = 1$. Let $\omega \in \Omega_0$ be an arbitrary sample point; from this point until the end of the argument, all quantities may be regarded as deterministic sequences evaluated at that ω.

Suppose, for contradiction, that $V_\infty(\omega) > 0$. There exist $0 < a < b < \infty$ and n_0 such that $a \leq V_n(\omega) \leq b$ for all $n \geq n_0$. Hence $PX_n(\omega)$ eventually belongs to

$$K_{a,b} := \{z \in \mathcal{H}_P : a \leq \mathcal{L}(z) - \mathcal{L}(z^*) \leq b\}.$$

Coercivity of $\mathcal{L}|_{\mathcal{H}_P}$ makes $K_{a,b}$ compact. The gradient of the restricted function $\mathcal{L}|_{\mathcal{H}_P}$ is $P\nabla\mathcal{L}(z)$. By (26), $P\nabla\mathcal{L}(z) = \nabla\mathcal{L}(z)$, so a point of $\mathcal{H}_P$ is stationary for the restriction exactly when its full gradient vanishes. Strict convexity and differentiability therefore imply that z^* is the only point of $\mathcal{H}_P$ with zero gradient. Since $z^* \notin K_{a,b}$, the continuous function $z \mapsto \|\nabla\mathcal{L}(z)\|$ is strictly positive on $K_{a,b}$; the extreme-value theorem therefore gives

$$\delta := \min_{z \in K_{a,b}} \|\nabla\mathcal{L}(z)\| > 0.$$

Equation (28) would then force $\sum_n \underline{\lambda}_n < \infty$, contradicting (16). Thus $V_n \to 0$ almost surely.

It remains only to convert loss convergence into convergence of the projected iterates. Fix $\varepsilon > 0$ and define

$$K_\varepsilon := \{z \in \mathcal{H}_P : \|z - z^*\| \geq \varepsilon, \quad \mathcal{L}(z) \leq \mathcal{L}(z^*) + 1\}.$$

Coercivity makes the sublevel set $\{z \in \mathcal{H}_P : \mathcal{L}(z) \leq \mathcal{L}(z^*) + 1\}$ compact, so K_ε is compact. If $K_\varepsilon = \varnothing$, then $V_n < 1$ already implies $\|PX_n - z^*\| < \varepsilon$ eventually. Hence assume $K_\varepsilon \neq \varnothing$ and set

$$\delta_\varepsilon := \min_{z \in K_\varepsilon} \big(\mathcal{L}(z) - \mathcal{L}(z^*)\big).$$

Continuity gives attainment of the minimum, and uniqueness of z^* on $\mathcal{H}_P$ gives $\delta_\varepsilon > 0$. Since $V_n \to 0$, eventually both $V_n < 1$ and $V_n < \delta_\varepsilon$. The first inequality places PX_n in the relevant compact sublevel set, while the second rules out $PX_n \in K_\varepsilon$. Hence eventually $\|PX_n - z^*\| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $PX_n \to z^*$ almost surely. □

C.2 LIFTING CONVERGENCE ALONG THE NEUTRAL SYMMETRY DIRECTION

This subsection contains the generic lifting criterion and the CElo-specific participation-clock argument alluded to in Section 3.4. These results strengthen the identifiable convergence of Corollary 1.

Theorem 3 (Convergence of a linear component under summable drift)**.** *Consider* (13) *and define*

$$h_n := \mathbb{E}[H_{n+1} \mid \mathcal{F}_n], \qquad \xi_{n+1} := H_{n+1} - h_n.$$

Let $B : \mathbb{R}^d \to \mathbb{R}^k$ *be a fixed linear map. Suppose, almost surely,*

$$\sum_{n=0}^{\infty} \|B\Lambda_n h_n\| < \infty \tag{29}$$

and

$$\sum_{n=0}^{\infty} \mathbb{E}\Big[\|B\Lambda_n \xi_{n+1}\|^2 \mid \mathcal{F}_n\Big] < \infty. \tag{30}$$

Then BX_n *converges almost surely to a finite random vector.*

In particular, (30) *follows if*

$$\mathbb{E}[\|\xi_{n+1}\|^2 \mid \mathcal{F}_n] \le C \quad \text{a.s.}$$

for a deterministic $C < \infty$ *and* $\sum_n \|B\Lambda_n\|_{\mathrm{op}}^2 < \infty$ *almost surely.*

Proof. Applying B to (13) and summing gives

$$BX_n = BX_0 + D_n + M_n,$$

where

$$D_n := \sum_{j=0}^{n-1} B\Lambda_j h_j, \qquad M_n := \sum_{j=0}^{n-1} B\Lambda_j \xi_{j+1}.$$

Condition (29) says exactly that the vector series

$$\sum_{j=0}^{\infty} B\Lambda_j h_j$$

is absolutely summable in the finite-dimensional normed space $\mathbb{R}^k$. Hence its partial sums D_n form a Cauchy sequence and converge to a finite vector.

The process (M_n) is a finite-dimensional locally square-integrable martingale, with increments

$$\Delta M_{n+1} = B\Lambda_n \xi_{n+1}.$$

For each scalar component $M^{(j)}$, condition (30) implies

$$\langle M^{(j)} \rangle_\infty := \sum_{n=0}^{\infty} \mathbb{E}\Big[(\Delta M_{n+1}^{(j)})^2 \mid \mathcal{F}_n\Big] < \infty \quad \text{a.s.}$$

Hall and Heyde (Hall & Heyde, 1980, Theorem 2.15), applied after the usual localisation, therefore gives

$$M_n^{(j)} \longrightarrow M_\infty^{(j)} \quad \text{a.s.}$$

for every component j. Since there are finitely many components, $M_n \to M_\infty$ almost surely in $\mathbb{R}^k$. Combining the limits of D_n and M_n proves convergence of BX_n. □

Lemma 2 (Asymptotic cancellation along a one-dimensional symmetry)**.** *Let*

$$h_n := \mathbb{E}[H_{n+1} \mid \mathcal{F}_n],$$

and suppose there is a fixed nonzero vector $u \in \mathbb{R}^d$ *such that*

$$u^\top h_n = 0 \qquad \textit{a.s. for every } n. \tag{31}$$

Let $w \in \mathbb{R}^d$ be fixed. If there exists an $\mathcal{F}_n$-measurable scalar sequence (a_n) such that

$$\sum_{n=0}^{\infty} \left\|\Lambda_n^\top w - a_n u\right\| \, \|h_n\| < \infty \qquad \text{a.s.}, \tag{32}$$

then

$$\sum_{n=0}^{\infty} \left|w^\top \Lambda_n h_n\right| < \infty \qquad \text{a.s.} \tag{33}$$

In particular, if $\sup_n \|h_n\| < \infty$ almost surely, it is enough that

$$\sum_{n} \left\|\Lambda_n^\top w - a_n u\right\| < \infty \qquad \text{a.s.}$$

Proof. For every n,

$$\begin{aligned} w^\top \Lambda_n h_n &= (\Lambda_n^\top w)^\top h_n \\ &= (\Lambda_n^\top w - a_n u)^\top h_n + a_n u^\top h_n \\ &= (\Lambda_n^\top w - a_n u)^\top h_n, \end{aligned}$$

where the last equality uses (31). Hence Cauchy–Schwarz gives

$$|w^\top \Lambda_n h_n| \le \left\|\Lambda_n^\top w - a_n u\right\| \, \|h_n\| .$$

Summing and applying (32) proves the claim. □

Lemma 3 (Lifting quotient convergence to point convergence)**.** *Let $u \in \mathbb{R}^d \setminus \{0\}$ and let P be the orthogonal projection onto $u^\perp$. Suppose*

$$PX_n \longrightarrow z^*, \qquad w^\top X_n \longrightarrow S_\infty$$

almost surely for some fixed $w \in \mathbb{R}^d$ with $w^\top u \neq 0$. Then there is a finite random scalar C_∞ such that

$$X_n \longrightarrow z^* + C_\infty u \qquad \textit{almost surely}.$$

More precisely,

$$C_\infty = \frac{S_\infty - w^\top z^*}{w^\top u}.$$

Proof. Because P projects onto $u^\perp$, there is a unique scalar C_n such that

$$X_n = PX_n + C_n u.$$

Applying $w^\top$ gives

$$C_n = \frac{w^\top X_n - w^\top P X_n}{w^\top u}.$$

The numerator converges by the two hypotheses and continuity of the fixed linear functional $w^\top$, while the denominator is nonzero. Thus $C_n \to C_\infty$, and substituting into the decomposition yields the stated pointwise limit. □

Lemma 4 (Mean field and bounded innovation)**.** *For the model of Section 2,*

$$\mathbb{E}[U_{n+1} \mid \mathcal{F}_n] = -\nabla \mathcal{L}_q(R_n) \quad \text{a.s.}, \tag{34}$$

and

$$\|U_{n+1}\|^2 \le 2 \quad \text{a.s.} \tag{35}$$

Proof. The mean-field identity (34) is verified in Appendix B; see (24)–(25). It remains to verify the uniform bound.

By definition,

$$\|U_{n+1}\|^2 = \left|Y_{n+1} - F(m_{E_{n+1}}^\top R_n)\right|^2 \left\|m_{E_{n+1}}\right\|^2 .$$

Both Y_{n+1} and $F(\cdot)$ lie in $[0,1]$, so the squared scalar factor is at most one. If $e = (i,j)$, then $m_e = \hat{e}_i - \hat{e}_j$, and hence

$$\|m_e\|^2 = \|\hat{e}_i - \hat{e}_j\|^2 = 2.$$

Therefore $\|U_{n+1}\|^2 \le 2$ almost surely. □

Lemma 5 (Geometry of the population loss)**.** *Let*

$$L_G(q) := \sum_{e \in E} q_e m_e m_e^\top$$

be the weighted graph Laplacian associated with the chosen edge orientations. Then:

(a) $\mathcal{L}_q(r + t\mathbf{1}) = \mathcal{L}_q(r)$ *for every* r *and* t*, and*

$$\nabla \mathcal{L}_q(r) = \nabla \mathcal{L}_q(\Pi r) \in \mathcal{H}.$$

(b) $\mathcal{L}_q$ *has globally Lipschitz gradient with smoothness constant*

$$L_q^{\mathrm{sm}} := \frac{1}{4} \lambda_{\max}(L_G(q)). \tag{36}$$

(c) $\mathcal{L}_q|_{\mathcal{H}}$ *is strictly convex.*

(d) $\mathcal{L}_q|_{\mathcal{H}}$ *is coercive. Consequently there is a unique*

$$z_q^* := \arg\min_{z \in \mathcal{H}} \mathcal{L}_q(z). \tag{37}$$

Proof. Because $m_e^\top \mathbf{1} = 0$, every term in (7) depends only on rating differences, proving translation invariance. Differentiating the identity $\mathcal{L}_q = \mathcal{L}_q \circ \Pi$ gives the gradient assertions as in (26).

The Hessian is

$$\nabla^2 \mathcal{L}_q(r) = \sum_{e \in E} q_e F'(m_e^\top r) m_e m_e^\top. \tag{38}$$

For every vector v, $m_e m_e^\top$ is positive semidefinite because

$$v^\top m_e m_e^\top v = (m_e^\top v)^2 \geq 0.$$

Since $0 < F'(t) \leq 1/4$, each summand in (38) satisfies

$$0 \preceq q_e F'(m_e^\top r) m_e m_e^\top \preceq \frac{q_e}{4} m_e m_e^\top.$$

Loewner inequalities are preserved under finite sums, so

$$0 \preceq \nabla^2 \mathcal{L}_q(r) \preceq \frac{1}{4} L_G(q).$$

Taking the largest Rayleigh quotient gives $\left\|\nabla^2 \mathcal{L}_q(r)\right\|_{\mathrm{op}} \leq \frac{1}{4} \lambda_{\max}(L_G(q))$ uniformly in r, which is equivalent to global Lipschitz continuity of the gradient with constant (36).

For $v \in \mathcal{H}$,

$$v^\top \nabla^2 \mathcal{L}_q(r) v = \sum_{e \in E} q_e F'(m_e^\top r)(m_e^\top v)^2.$$

If this is zero then every nonnegative summand must vanish, so $m_e^\top v = 0$ for every edge e. For an oriented edge $e = (i, j)$ this condition is simply

$$v_i - v_j = 0,$$

so the two endpoint coordinates agree. If i and j are arbitrary vertices, connectedness supplies a path $i = v_0, v_1, \ldots, v_k = j$. Equality across every edge of that path gives

$$v_i = v_{v_0} = v_{v_1} = \cdots = v_{v_k} = v_j.$$

Thus every coordinate of v has the same value: $v = t\mathbf{1}$. Since $v \in \mathcal{H} = \mathbf{1}^\perp$, one has $0 = \mathbf{1}^\top v = Nt$, hence $t = 0$.

Consequently, for every nonzero direction $v \in \mathcal{H}$,

$$v^\top \nabla^2 \mathcal{L}_q(r) v > 0.$$

This is precisely what is meant by saying that the Hessian is positive definite *on the subspace* $\mathcal{H}$. It need not be positive definite on all of $\mathbb{R}^N$: indeed the translation direction $\mathbf{1}$ is a null direction. Positive definiteness along every nonzero direction in $\mathcal{H}$ implies strict convexity of $\mathcal{L}_q|_{\mathcal{H}}$.

For coercivity, define

$$\varepsilon := \min_{e\in E} \min\{p_e, 1-p_e\} > 0, \qquad q_{\min} := \min_{e\in E} q_e > 0.$$

For every $p \in (0,1)$ and $d \in \mathbb{R}$,

$$\begin{aligned} -p\log F(d) - (1-p)\log(1-F(d)) &= p\log(1+e^{-d}) + (1-p)\log(1+e^{d}) \\ &\geq \min\{p, 1-p\}|d|. \end{aligned} \tag{39}$$

Indeed, if $d \geq 0$ retain only the second term and use $\log(1+e^d) \geq d$; if $d \leq 0$ retain only the first term and use $\log(1+e^{-d}) \geq -d$.

Let

$$\Delta(r) := \max_i r_i - \min_i r_i.$$

If $\Delta(r) = 0$, the lower bound below is immediate; moreover, $r \in \mathcal{H}$ then forces $r = 0$. Hence assume $\Delta(r) > 0$. Choose vertices v_0 and v_k attaining the maximum and minimum respectively, and take a simple path

$$v_0, v_1, \ldots, v_k$$

between them. A simple path visits no vertex twice, so $k \leq N-1$. Along this path the endpoint difference telescopes:

$$\begin{aligned} \Delta(r) = r_{v_0} - r_{v_k} &= \sum_{\ell=0}^{k-1} (r_{v_\ell} - r_{v_{\ell+1}}) \\ &\leq \sum_{\ell=0}^{k-1} |r_{v_\ell} - r_{v_{\ell+1}}|. \end{aligned}$$

If every one of the k edge differences were smaller than $\Delta(r)/k$, the final sum would be smaller than $\Delta(r)$, a contradiction. Thus at least one path edge e satisfies

$$|m_e^\top r| \geq \frac{\Delta(r)}{k} \geq \frac{\Delta(r)}{N-1}. \tag{40}$$

All terms of (7) are nonnegative, so (39) and (40) give

$$\mathcal{L}_q(r) \geq \frac{q_{\min}\varepsilon}{N-1}\Delta(r). \tag{41}$$

For $r \in \mathcal{H}$, zero lies between $\min_i r_i$ and $\max_i r_i$, hence $|r_i| \leq \Delta(r)$ for every i and therefore

$$\Delta(r) \geq \frac{\|r\|}{\sqrt{N}}. \tag{42}$$

Combining (41) and (42) proves coercivity on $\mathcal{H}$. Existence and uniqueness of (37) follow from coercivity, continuity, and strict convexity. □

Proof of Proposition 1. For $s \in (0,1)$,

$$-p\log s - (1-p)\log(1-s) = H_{\mathrm{B}}(p) + D_{\mathrm{KL}}(\mathrm{Bern}(p)\|\mathrm{Bern}(s)),$$

so summing edgewise proves (11). Lemma 5 then makes the minimiser on $\mathcal{H}$ unique.

For (i), $\Pi\theta$ has the same edge differences as θ, because $m_e^\top \mathbf{1} = 0$. Hence every KL term in (11) is zero at $\Pi\theta$. Since KL divergence is nonnegative, this is a global minimum for every positive choice of the weights q_e, and uniqueness on $\mathcal{H}$ gives $z_q^* = \Pi\theta$.

For (ii), it suffices to give a counterexample. Take the complete graph on three vertices, orient the edges as $(1,2)$, $(2,3)$ and $(1,3)$, and set

$$p_{12} = 0.9, \qquad p_{23} = 0.9, \qquad p_{13} = 0.1.$$

These probabilities are not Bradley–Terry realisable, since their logits would require simultaneously

$$\theta_1 - \theta_2 > 0, \qquad \theta_2 - \theta_3 > 0, \qquad \theta_1 - \theta_3 < 0.$$

For uniform matchmaking $q = (1/3, 1/3, 1/3)$, evaluating (24) at $r = 0$ gives

$$\nabla \mathcal{L}_q(0) = \frac{1}{3}\big[-0.4(e_1 - e_2) - 0.4(e_2 - e_3) + 0.4(e_1 - e_3)\big] = 0.$$

Strict convexity on $\mathcal{H}$ therefore gives $z_q^* = 0$. Now change only the matchmaking weights to

$$q'_{12} = 0.45, \qquad q'_{23} = 0.45, \qquad q'_{13} = 0.10.$$

Then

$$\nabla \mathcal{L}_{q'}(0) = -0.14e_1 + 0.14e_3 \neq 0.$$

By Lemma 5(a), this gradient lies in $\mathcal{H}$, so the first-order condition for minimisation on $\mathcal{H}$ fails at 0. Thus $z_{q'}^* \neq z_q^*$, proving that the pseudo-true target can depend on matchmaking under misspecification. □

Lemma 6 (Power-law participation clocks satisfy the SA step conditions)**.** *For each player define the participation probability*

$$\pi_i := \sum_{e \ni i} q_e. \tag{43}$$

Then, for every match index k, $\pi_i = \mathbb{P}(i \in E_k)$; in particular $\pi_i > 0$ for every i. For the step sizes (4) with $1/2 < \alpha \leq 1$,

$$\sum_n \underline{\lambda}_n = \infty, \qquad \sum_n \overline{\lambda}_n^2 < \infty \quad \textit{a.s.} \tag{44}$$

Proof. Connectedness and $q_e > 0$ give $\pi_i > 0$. Since $\nu_i(n) \leq n$,

$$\lambda_{n,i} \geq \frac{c_{\min}}{(a_{\max} + n)^\alpha},$$

where $c_{\min} := \min_i c_i$ and $a_{\max} := \max_i a_i$. Hence

$$\underline{\lambda}_n \geq \frac{c_{\min}}{(a_{\max} + n)^\alpha},$$

and the first series in (44) diverges because $\alpha \leq 1$.

For each fixed i, the indicators $\mathbf{1}\{i \in E_k\}$ are i.i.d. Bernoulli(π_i), so the strong law gives

$$\frac{\nu_i(n)}{n} \longrightarrow \pi_i \quad \text{a.s.}$$

On the probability-one event where this convergence holds for player i, there is a finite random index N_i such that for all $n \geq N_i$,

$$\left|\frac{\nu_i(n)}{n} - \pi_i\right| < \frac{\pi_i}{2},$$

hence $\nu_i(n) \geq (\pi_i/2)n$. There are only finitely many players, so intersecting their probability-one events still gives a probability-one event, and taking $N_0 := \max_i N_i$ makes the lower bound hold for every player simultaneously for all $n \geq N_0$.

Consequently, for $n \geq N_0$,

$$\lambda_{n,i}^2 = \frac{c_i^2}{(a_i + \nu_i(n))^{2\alpha}} \leq c_i^2 \left(\frac{2}{\pi_i}\right)^{2\alpha} n^{-2\alpha}.$$

Since there are finitely many players, the deterministic constant

$$C := \max_i c_i^2 \left(\frac{2}{\pi_i}\right)^{2\alpha} < \infty$$

satisfies, eventually almost surely,

$$\overline{\lambda}_n^2 \leq C n^{-2\alpha}.$$

The second series in (44) converges because $2\alpha > 1$. □

Proof of Corollary 1. Apply Theorem 2 with $X_n = R_n$, $H_{n+1} = U_{n+1}$, $P = \Pi$, and $\mathcal{L} = \mathcal{L}_q$. Lemma 4 gives the negative-gradient mean condition and the uniform bound $\mathbb{E}[\|U_{n+1}\|^2 \mid \mathcal{F}_n] \le 2$, so (17) holds with $A = 2$ and $B = 0$. Lemma 5 supplies smoothness, invariance, strict convexity and coercivity on $\mathcal{H}$, Lemma 6 supplies (16), and (22) supplies the required initial integrability. □

For the CElo lifting argument, recall the participation probabilities π_i from (43) and define

$$w_i := \frac{\pi_i^\alpha}{c_i} > 0, \qquad i = 1, \dots, N. \tag{45}$$

The strong law gives $\nu_i(n)/n \to \pi_i$, so $\lambda_{n,i} w_i \approx n^{-\alpha}$ at first order. Pointwise alignment is not enough: the accumulated mismatch must be finite.

Lemma 7 (Participation-clock balancing)**.** *For w defined in (45), there exists a random remainder sequence $(r_n)_{n\ge1}$ such that, for $n \ge 1$,*

$$\Lambda_n w = n^{-\alpha}\mathbf{1} + r_n, \qquad \sum_{n=1}^{\infty} \|r_n\| < \infty \qquad \textit{almost surely}. \tag{46}$$

Proof of Lemma 7. For fixed i, write

$$I_{k,i} := \mathbf{1}\{i \in E_k\}, \qquad \nu_i(n) = \sum_{k=1}^{n} I_{k,i}.$$

The variables $(I_{k,i})_{k\ge1}$ are i.i.d. Bernoulli(π_i). Hoeffding's inequality gives, for every $t > 0$,

$$\mathbb{P}(|\nu_i(n) - \pi_i n| \ge t) \le 2\exp\left(-\frac{2t^2}{n}\right).$$

Choose $t = K\sqrt{n \log n}$ with any $K > 1/\sqrt{2}$. Then

$$\mathbb{P}\Big(|\nu_i(n) - \pi_i n| > K\sqrt{n \log n}\Big) \le 2n^{-2K^2},$$

and the right-hand side is summable in n. By the first Borel–Cantelli lemma, almost surely only finitely many of these deviation events occur. Thus

$$\nu_i(n) - \pi_i n = O(\sqrt{n \log n}) \qquad \text{a.s.} \tag{47}$$

Because there are finitely many players, the rate may be taken simultaneously in i on one probability-one event.

By Lemma 6, the strong law also yields $a_i + \nu_i(n) \ge (\pi_i/2)n$ eventually. Applying the mean value theorem to $f(x) = x^{-\alpha}$ between $a_i + \nu_i(n)$ and $\pi_i n$, and using this linear lower bound on the intermediate point, gives

$$\begin{aligned}\left|\frac{\pi_i^\alpha}{(a_i + \nu_i(n))^\alpha} - \frac{1}{n^\alpha}\right| &\le C_i n^{-\alpha-1} \left|a_i + \nu_i(n) - \pi_i n\right| \\ &= O\left(\frac{\sqrt{\log n}}{n^{\alpha+1/2}}\right) + O(n^{-\alpha-1}) \quad \text{a.s.}\end{aligned} \tag{48}$$

Since $\alpha > 1/2$, both deterministic comparison series are summable. There are finitely many coordinates, so (48) implies

$$\sum_{n=1}^{\infty} \left\|\Lambda_n w - n^{-\alpha}\mathbf{1}\right\| < \infty \quad \text{a.s.}$$

Defining $r_n := \Lambda_n w - n^{-\alpha}\mathbf{1}$ proves (46). □

Corollary 2 (Martingale control for CElo)**.** *For w defined in (45),*

$$\sum_{n=0}^{\infty} \mathbb{E}\Big[\left|w^\top \Lambda_n \left(U_{n+1} + \nabla\mathcal{L}_q(R_n)\right)\right|^2 \Big| \mathcal{F}_n\Big] < \infty \qquad \textit{almost surely}. \tag{49}$$

Proof. Lemma 4 gives

$$\mathbb{E}[U_{n+1} + \nabla \mathcal{L}_q(R_n) \mid \mathcal{F}_n] = 0.$$

Since conditional centring does not increase the conditional second moment,

$$\mathbb{E}\Big[\|U_{n+1} + \nabla \mathcal{L}_q(R_n)\|^2 \Big| \mathcal{F}_n\Big] \le \mathbb{E}[\|U_{n+1}\|^2 \mid \mathcal{F}_n] \le 2.$$

Therefore

$$\begin{aligned}\mathbb{E}\Big[\big|w^\top \Lambda_n (U_{n+1} + \nabla \mathcal{L}_q(R_n))\big|^2 \Big| \mathcal{F}_n\Big] &\le \|\Lambda_n w\|^2 \, \mathbb{E}\Big[\|U_{n+1} + \nabla \mathcal{L}_q(R_n)\|^2 \Big| \mathcal{F}_n\Big] \\ &\le 2 \|w\|^2 \overline{\lambda}_n^2.\end{aligned}$$

Lemma 6 makes the final series summable almost surely, proving (49). □

Proof of Theorem 1. Set

$$h_n := \mathbb{E}[U_{n+1} \mid \mathcal{F}_n] = -\nabla \mathcal{L}_q(R_n).$$

Translation invariance gives

$$\mathbf{1}^\top h_n = 0. \tag{50}$$

Also, from (24),

$$|\nabla_i \mathcal{L}_q(r)| \le \sum_{e \ni i} q_e = \pi_i \le 1, \tag{51}$$

so $\sup_n \|h_n\| \le \sqrt{N}$.

Apply Lemma 2 with

$$u = \mathbf{1}, \qquad a_0 := 0, \quad a_n := n^{-\alpha} \text{ for } n \ge 1, \qquad w_i = \frac{\pi_i^\alpha}{c_i}.$$

Lemma 7 supplies the summability of $\|\Lambda_n w - n^{-\alpha}\mathbf{1}\|$ for $n \ge 1$. Together with the single finite $n = 0$ term and $\Lambda_n^\top = \Lambda_n$, this gives

$$\sum_{n=0}^{\infty} \big\|\Lambda_n^\top w - a_n u\big\| < \infty \quad \text{a.s.}$$

Therefore

$$\sum_n |w^\top \Lambda_n h_n| < \infty \quad \text{a.s.}$$

Corollary 2 supplies the martingale condition of Theorem 3. Applying that theorem with $B = w^\top$ yields

$$w^\top R_n \longrightarrow S_\infty \qquad \text{a.s.} \tag{52}$$

for a finite random scalar S_∞.

Corollary 1 gives $\Pi R_n \to z_q^*$ almost surely. Since Π is the orthogonal projection onto $\mathbf{1}^\perp$ and

$$w^\top \mathbf{1} = \sum_i \frac{\pi_i^\alpha}{c_i} > 0,$$

Lemma 3 applies with $P = \Pi$ and $u = \mathbf{1}$. Hence

$$R_n \longrightarrow z_q^* + C_\infty \mathbf{1} \quad \text{a.s.}, \qquad C_\infty = \frac{S_\infty - w^\top z_q^*}{w^\top \mathbf{1}}.$$

□

D Algorithm Details

Algorithm 1 is a high-level algorithmic implementation of CElo. In this case, κ_i is identical to ν_i from Section 2 – this notation accommodates Algorithm 2's resetting behaviour.

Algorithm 1 Convergent Elo

Require: Initial ratings $R_0 \in \mathbb{R}^N$; $c > 0$, $M > 0$, $\alpha \in (1/2, 1]$

Initialise effective counts $\kappa_i \leftarrow 0$ for all $i \in \{1, \dots, N\}$
for $n = 0, 1, 2, \dots$ **do**
  Observe match $E_{n+1} = (i, j)$ and score $Y_{n+1} \in [0, 1]$ for player i
  $\widehat{p}_{n+1} \leftarrow F(R_{n,i} - R_{n,j})$
  $\varepsilon_{n+1} \leftarrow Y_{n+1} - \widehat{p}_{n+1}$
  $R_{n+1,i} \leftarrow R_{n,i} + \dfrac{c}{(M + \kappa_i)^\alpha} \varepsilon_{n+1}$
  $R_{n+1,j} \leftarrow R_{n,j} - \dfrac{c}{(M + \kappa_j)^\alpha} \varepsilon_{n+1}$
  $R_{n+1,\ell} \leftarrow R_{n,\ell}$ for all $\ell \notin \{i, j\}$
  $\kappa_i \leftarrow \kappa_i + 1$
  $\kappa_j \leftarrow \kappa_j + 1$
end for

D.1 Heuristic nonstationary extension

CElo's diminishing gains deliberately trade tracking responsiveness for convergence. CUSUM-CElo (shortened to C-CElo) is an exploratory engineering extension for the applied experiments. Its implementation is shown in Algorithm 2. C-CElo wraps the convergent CElo core in a two-sided CUSUM detector on signed prediction residuals (Page, 1954). Upon a detection, it partially resets the affected player's effective local count, temporarily increasing its gain while leaving the rating update itself unchanged. The recursive unanchored reset, reported below as *C-CElo (zero anchor)*, maps $\kappa_i \leftarrow \rho\kappa_i$; the lifetime-anchored variant, reported as *C-CElo*, also maintains the cumulative appearance count ν_i, which is never reset, and instead maps $\kappa_i \leftarrow \rho\nu_i$. Neither resetting process is covered by the stationary convergence theorem. Their role is to test whether responsiveness can be layered onto a convergent core when effective strengths change with time.

Algorithm 2 C-CElo (zero anchor) with soft effective-count resets

Require: Initial ratings $R_0 \in \mathbb{R}^N$; $c > 0$, $M > 0$, $\alpha \in (1/2, 1]$; CUSUM drift $\delta \geq 0$; threshold $h > 0$; reset factor $\rho \in (0, 1]$

Initialise $\kappa_i \leftarrow 0$, $C_i^+ \leftarrow 0$, and $C_i^- \leftarrow 0$ for all $i \in \{1, \dots, N\}$
for $n = 0, 1, 2, \dots$ **do**
  Observe match $E_{n+1} = (i, j)$ and score $Y_{n+1} \in [0, 1]$ for player i
  $\widehat{p}_{n+1} \leftarrow F(R_{n,i} - R_{n,j})$
  $\varepsilon_i \leftarrow Y_{n+1} - \widehat{p}_{n+1}$
  $\varepsilon_j \leftarrow -\varepsilon_i$
  for $\ell \in \{i, j\}$ **do**
    $R_{n+1,\ell} \leftarrow R_{n,\ell} + \dfrac{c}{(M + \kappa_\ell)^\alpha} \varepsilon_\ell$
    $C_\ell^+ \leftarrow \max\{0,\, C_\ell^+ + \varepsilon_\ell - \delta\}$
    $C_\ell^- \leftarrow \max\{0,\, C_\ell^- - \varepsilon_\ell - \delta\}$
    if $C_\ell^+ > h$ **or** $C_\ell^- > h$ **then**
      $\kappa_\ell \leftarrow \rho\kappa_\ell$
      $C_\ell^+ \leftarrow 0$
      $C_\ell^- \leftarrow 0$
    end if
    $\kappa_\ell \leftarrow \kappa_\ell + 1$
  end for
  $R_{n+1,\ell} \leftarrow R_{n,\ell}$ for all $\ell \notin \{i, j\}$
end for

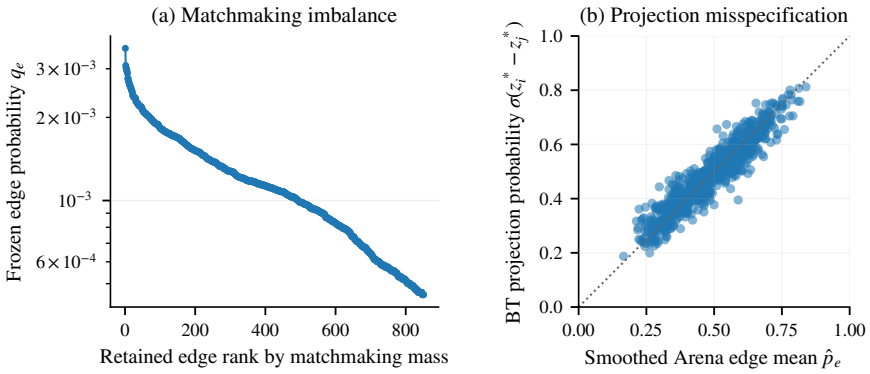

Figure 3: Diagnostics for the Arena-derived stationary population. Panel (a) sorts the 851 retained edge probabilities q_e, showing substantial matchmaking imbalance even after the low-count filter. Panel (b) compares the smoothed edge means $\widehat{p}_e$ with probabilities implied by the single q-weighted pseudo-true Bradley–Terry target; departures from the diagonal visualise the misspecification that makes the target a genuine projection rather than a recovered latent-strength vector.

E Experimental Details

E.1 Arena-derived stationary settling experiment

Population construction. The stationary resampling experiment uses the same Arena Text source window as the applied evaluation: April 1 through July 31, 2025. We remove "both bad" outcomes, retain all languages and both code and non-code prompts, and encode an ordinary tie as score $1/2$ when estimating an edge mean. After aggregating unordered model pairs, we require at least 50 observations per pair and retain the largest connected component. This leaves $N = 50$ models, $E = 851$ supported pairs, and 109,646 retained source comparisons. Edge counts range from 50 to 390, with median 121.

For each retained edge e, the frozen matchmaking probability is its empirical share of retained comparisons,

$$q_e = \frac{n_e}{\sum_f n_f}.$$

Writing S_e for the sum of oriented scores on that edge, we use the symmetric Beta–Bernoulli smoothing rule

$$\widehat{p}_e = \frac{S_e + 1/2}{n_e + 1},$$

which keeps every frozen mean strictly between zero and one. The resulting values range from approximately 0.166 to 0.839. The target z_q^* is computed before simulation by minimising (7) after fixing one reference coordinate and then recentring the solution. The numerical optimisation uses L-BFGS-B with primary settings `gtol`=10^{-11}, `ftol`=10^{-15}, at most 2,000 iterations, and at most 80 line-search steps; a failed solve is retried from zero with `gtol`=10^{-10}, at most 4,000 iterations, and at most 100 line-search steps. The centred gradient norm at the returned target is 2.26×10^{-9}. The q-weighted Bernoulli-KL gap at the optimum is 3.33×10^{-3}, so the frozen edge means are not exactly represented by a single Bradley–Terry vector at numerical precision.

Online simulation. Each replicate draws a fresh i.i.d. stream from the frozen population,

$$E_{n+1} \sim q, \qquad Y_{n+1} \mid E_{n+1} = e \sim \text{Bernoulli}(\widehat{p}_e),$$

and every method receives the same sampled edge and outcome sequence within a replicate. We run 32 independent replicates to a horizon of 10^8 comparisons. The continuation beyond the 109,646 retained source comparisons is therefore extrapolation of the fixed empirical population, not additional Arena observation. This Bernoulli resampling preserves the smoothed edge means and empirical matchmaking frequencies, but not the source draw distribution or conditional outcome

variance; the finite-horizon fluctuations and fixed-K error floor should therefore be interpreted as results for this synthetic noise model.

Here, CElo gains are parameterised as

$$\eta_{n,i}(\alpha) = \eta_0 \left(\frac{M}{M + \nu_i(n)} \right)^{\alpha},$$

with $\eta_0 = 0.030469$ and $M = 70.2126$ held fixed across $\alpha \in \{0.51, 0.60, 0.75, 0.90, 1.00\}$. The value of M is derived from the prequential Arena Text validation-selected CElo configuration $(\eta_0, \alpha, \text{half-life}) = (0.030469, 0.58054, 161.5)$ re-parameterised into the form shown; no α is re-tuned on the settling experiment. Fixed-K Elo uses $K = 0.0066827$, also taken from the Arena Text validation-selected configuration rather than re-tuned on the frozen population.

Figure 1 reports coordinate-centred states. At 227 log-spaced checkpoints we compute $\|\Pi R_n - z_q^*\|_2$ and, when defined, the trailing-window movement $\|\Pi R_n - \Pi R_{n-W}\|_2$ with $W = 10{,}000$. Curves are medians and bands are 20–80% cross-stream ranges. These are Monte Carlo trajectory summaries, not confidence intervals for a finite-time theorem, and no convergence-rate slope is fit.

E.2 Controlled validation of unrecentred convergence

This experiment aims to empirically illustrate the convergent dynamics implied by Theorem 1. The experiment uses a complete graph of ten participants, with centred Bradley–Terry strengths given by ten evenly spaced values from -1.2 to 1.2. Each regime contains ten independent streams of 100,000 comparisons. CElo uses $c = 1.5$, $M = 8$, and $\alpha = 0.75$, so its pre-match gain for player i is $1.5/(8 + \nu_i)^{0.75}$. Fixed-K Elo uses $K = 0.1$ on the same natural logistic-rating scale; this value is used to illustrate the qualitative fixed-gain behaviour rather than as a tuned benchmark.

For homogeneous participation, every complete-graph edge is sampled uniformly. For heterogeneous participation, participant activity weights are

$$(1.00, 1.00, 0.60, 0.50, 0.30, 0.15, 0.1, 0.05, 0.03, 0.01),$$

and the probability of edge (i, j) is proportional to the product of the two endpoint activities. This gives participation probabilities approximately

$$(0.4870, 0.4870, 0.3349, 0.2880, 0.1834, 0.0957, 0.0647, 0.0328, 0.0198, 0.0066),$$

while retaining positive mass on every edge. CElo and fixed-K Elo are run on exactly the same sampled comparison stream within each replicate.

Figure 4 distinguishes centred convergence from convergence of the raw common offset. The wide late-time band in panel (c) reflects dispersion across streams approaching different finite offsets, rather than persistent fluctuations within individual streams; their median remains close to zero. This is consistent with Theorem 1, which guarantees point convergence along each sample path while allowing the limiting offset C_∞ to differ between paths.

E.3 Synthetic matchmaking experiment

Here we illustrate the q-dependence seen in Proposition 1 experimentally. The oriented edges are $(1, 2)$, $(2, 3)$, and $(1, 3)$ with outcome means $(0.9, 0.9, 0.1)$. The matchmaking distribution is parameterised as

$$q(\tau) = (1 - \tau)(0.10, 0.10, 0.80) + \tau(0.45, 0.45, 0.10), \qquad \tau \in [0, 1],$$

with uniform matchmaking achieved at $\tau = 2/3$. The solid curves are numerical minimisers of $\mathcal{L}_q$ over the centred subspace, evaluated on 200 values of τ. The realisable control uses centred strengths $(1.0, 0, -1.0)$ and the same matchmaking path.

Figure 5 traces the complete q-weighted KL target as matchmaking changes and overlays finite CElo endpoints. Along this path, changing only comparison frequencies reverses players 1 and 3 in the misspecified system, whereas the Bradley–Terry-realisable target remains invariant. This is the distinction in Proposition 1: under misspecification the comparison policy is part of the population estimand, while under realisability positive matchmaking weights affect sampling rather than the

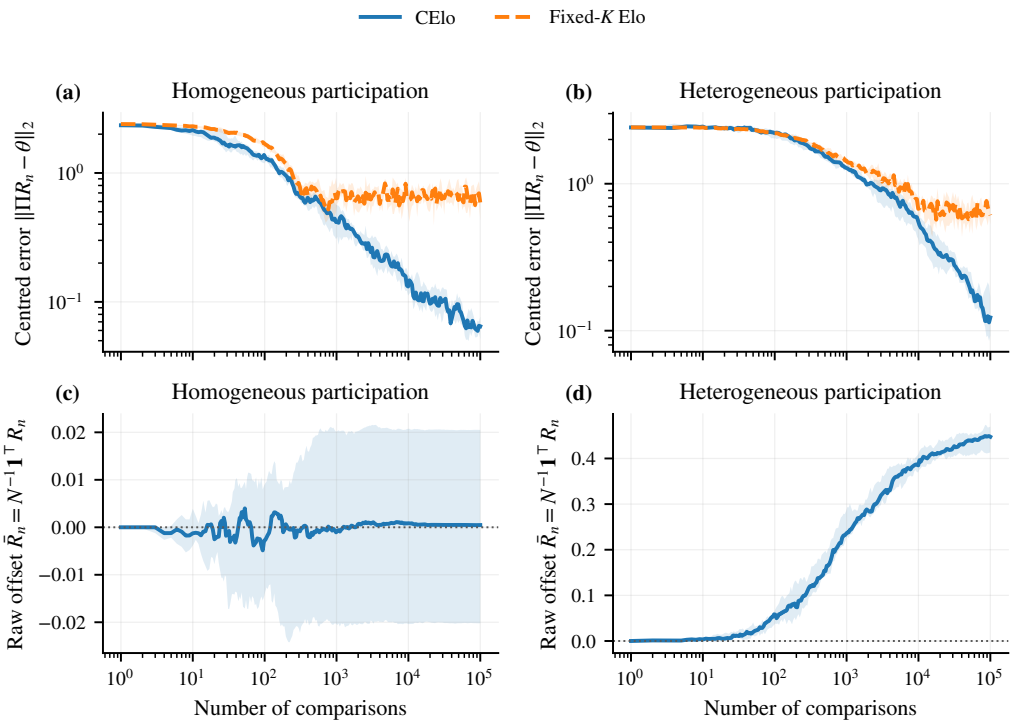

Figure 4: Controlled validation of identifiable and unrecentred convergence under homogeneous and heterogeneous participation. Top: centred target error (medians; 20–80% ranges across ten streams). Bottom: CElo's raw common offset. CElo settles while fixed-K Elo remains noisy; the late spread in panel (c) is across sample-path-specific limits. Raw-offset panels use different vertical scales.

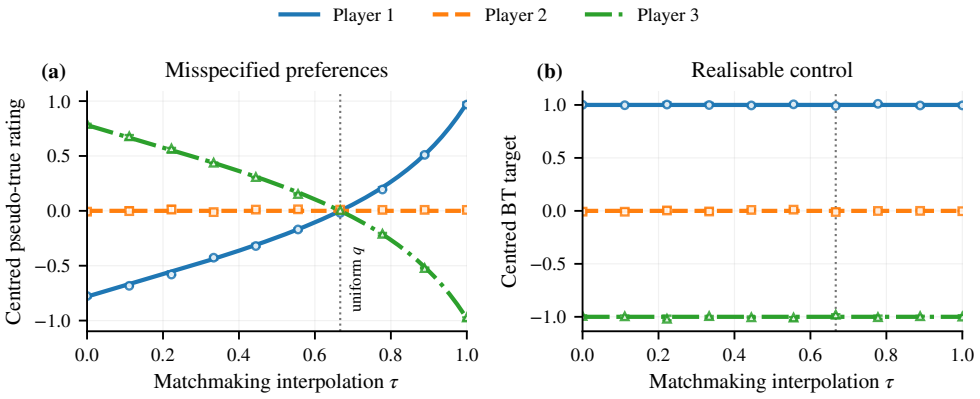

Figure 5: Illustration of the q-weighted pseudo-true target. Curves are exact centred targets; markers and bars show medians and 10–90% ranges of finite CElo runs. Matchmaking moves the misspecified projection, while the realisable target remains invariant; the dotted line is uniform matchmaking.

centred target. Qualitative matchmaking sensitivity under misspecification is not claimed to be unique to this work.

The finite-run overlays use CElo with $(c, M, \alpha) = (1.5, 8, 0.75)$. At ten values of τ, four independent streams of $50{,}000$ comparisons are simulated; markers are coordinate-wise medians and error bars are 10–90% ranges (even at these ranges, the error bars are too small to be immediately visible on the plots). These overlays are included only to connect the population optimiser charac-

terisation to the online recursion; the population curves themselves are the estimands described by Proposition 1.

E.4 Participant-local versus global clocks

Mechanism and experimental question. The ablation in Section 4 isolates the choice of clock within an otherwise identical diminishing-gain schedule. For participant i, let $\nu_i(n)$ count appearances in the first n comparisons. The participant-local and globally clocked gains used for comparison $n+1$ are

$$\lambda_{n,i}^{\mathrm{local}} = \eta_0 \left(\frac{M}{M + \nu_i(n)} \right)^{\alpha}, \qquad \lambda_n^{\mathrm{global}} = \eta_0 \left(\frac{M}{M + 2n/N} \right)^{\alpha}.$$

Every comparison contributes two participant appearances, so $2n/N$ is the average count per participant and equals the expected local count under homogeneous exposure. More generally, if π_i is participant i's marginal probability of appearing in a comparison, then $\mathbb{E}[\nu_i(n)] = n\pi_i$. Under the fixed i.i.d. matchmaking used here, the gain ratio satisfies

$$\frac{\lambda_n^{\mathrm{global}}}{\lambda_{n,i}^{\mathrm{local}}} = \left(\frac{M + \nu_i(n)}{M + 2n/N} \right)^{\alpha} \longrightarrow \left(\frac{N\pi_i}{2} \right)^{\alpha} \qquad \text{almost surely.}$$

Under homogeneous exposure the gain ratio tends to one, although realised local counts fluctuate at finite n. For rarely observed participants ($\pi_i < 2/N$), the global counter *overstates* their accumulated exposure and reduces their gain relative to the local schedule; for frequently observed participants ($\pi_i > 2/N$), it does the opposite. The experiment tests the finite-horizon consequences of this mismatch with a fixed conditional outcome distribution and population target, using identical realised comparisons within each paired local/global run.

Population and exposure regimes. We use $N = 50$ participants with a fixed, Bradley–Terry-realisable target. Fifty evenly spaced strengths on $[-1.35, 1.35]$ are randomly permuted and centred to obtain θ. Independently, each participant receives one activity score s_i from a permutation of 50 evenly spaced values on $[-0.5, 0.5]$. For a raw activity-weight ratio ρ, let $w_i = \rho^{s_i}$ and define the matchmaking policy and participant marginals by

$$q_{ij} = \frac{w_i w_j}{\sum_{a<b} w_a w_b} \quad (i < j), \qquad \pi_i = \sum_{j \neq i} q_{ij},$$

where $q_{ji} = q_{ij}$. We sample pairs by drawing two endpoints independently with probabilities proportional to w_i, rejecting and redrawing the entire pair whenever the endpoints coincide; the accepted unordered pair has exactly the distribution q. Conditional on the pair, outcomes follow $Y \mid (i,j) \sim \mathrm{Bernoulli}(F(\theta_i - \theta_j))$. Hence varying ρ changes only the allocation of observations, not the true pairwise probabilities or the centred target θ.

We consider $\rho \in \{1, \sqrt{10}, 10, \sqrt{1000}, 100\}$, giving actual marginal participation-imbalance ratios $\pi_{\max}/\pi_{\min}$ of 1.0000, 3.0888, 9.5410, 29.4705, and 91.0298, respectively. We report the marginal ratios, rather than the raw weight ratios, because π_i governs each participant's expected comparison count and therefore the clock mismatch.

Matched schedules and Monte Carlo protocol. All ratings start at zero. We use $\eta_0 = 0.030469$ and $M = 70.2126$ from the Arena prequential calibration and hold both fixed across methods and decay exponents. Equivalently, the two schedules above share $c_\alpha = \eta_0 M^\alpha$ in the form $c_\alpha/(M + \mathrm{clock})^\alpha$. The main comparison uses $\alpha = 0.75$ ($c_{0.75} = 0.739043$); the exponent check uses $\alpha \in \{0.60, 0.75, 1.00\}$, with corresponding $c_\alpha = 0.390578$, 0.739043, and 2.13931. For each fixed α, only the clock differs between the two algorithms, and both begin with gain η_0.

For each exposure regime we generate 32 independent paired streams of 10^7 comparisons. Within a stream, every method and tested exponent receives exactly the same sampled pairs and realised outcomes. We record 227 distinct, approximately logarithmically spaced checkpoints, including the first comparison and the endpoint. The primary trajectory metric is centred target error $\|\Pi R_n - \theta\|_2$. We also calculate the exact q-weighted population excess log-loss $\mathcal{L}_q(R_n) - \mathcal{L}_q(\theta)$ and record terminal participant-wise centred errors and realised participation counts. In particular, Figure 6(a) uses each participant's terminal root-mean-squared centred error across the 32 paired streams, rather than connecting different participants into a trajectory.

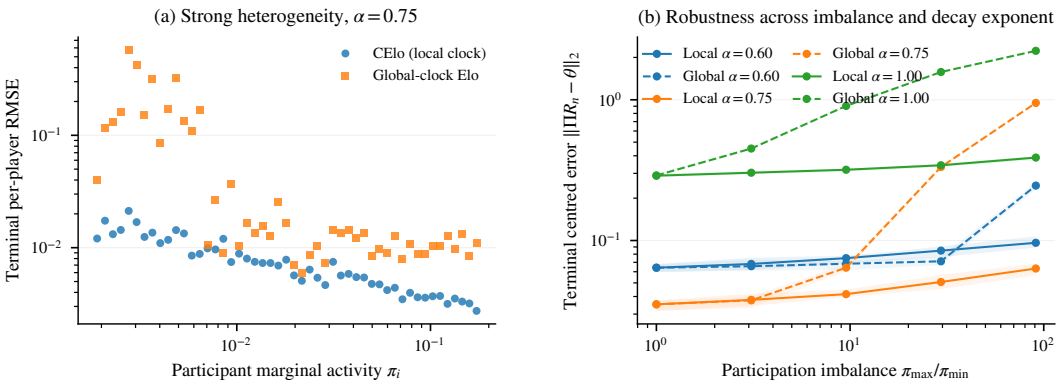

Figure 6: Participant-level and exposure-imbalance diagnostics for local and global clocks. Panel (a) plots terminal per-player RMSE of centred ratings across 32 paired streams against marginal participation probability π_i in the strongest-heterogeneity regime at $\alpha = 0.75$. Each marker represents a participant (circles: local clock; squares: global clock); participants are not connected by lines. Panel (b) reports median terminal centred target error over the five exposure regimes for local and global clocks with $\alpha \in \{0.60, 0.75, 1.00\}$; shaded regions are 20–80% ranges across paired streams. Its horizontal axis is the actual participation ratio $\pi_{\max}/\pi_{\min}$, not the raw activity-weight ratio.

Finite-horizon results and participant-level diagnostic. Figure 2 shows that the local and global schedules are nearly indistinguishable under homogeneous exposure: at $\alpha = 0.75$, their median terminal centred errors are 0.03519 and 0.03520, respectively. At mild imbalance ($\pi_{\max}/\pi_{\min} = 3.0888$), the corresponding values remain close (0.03786 and 0.03758). The difference grows with heterogeneity: at marginal ratios 9.5410 and 29.4705, the local/global errors are $0.04157/0.06409$ and $0.05072/0.3331$, respectively. In the strongest regime ($\pi_{\max}/\pi_{\min} = 91.0298$), they reach 0.06324 and 0.9502 after 10^7 comparisons. This progression is consistent with the increasing mismatch between individual exposure and the mean global clock.

Figure 6(a) identifies where the strongest-regime error gap arises. The largest discrepancies occur among participants with small π_i, whose gains are reduced prematurely by the global counter. For scale, at $n = 10^7$ the least active participant has about $19{,}000$ expected appearances, whereas the global counter is $2n/N = 400{,}000$; at $\alpha = 0.75$, substituting the expected local count gives a global gain approximately one tenth of the local gain. Among the lowest-activity quartile, the median across participants of their terminal, across-stream median absolute errors is 0.008892 locally versus 0.1505 globally. In the highest-activity quartile the corresponding values are 0.002343 and 0.008037. The discrepancy is concentrated among rare participants, rather than confined exclusively to them.

Sensitivity to exposure imbalance and decay exponent. Figure 6(b) extends the comparison to all five exposure regimes and three decay exponents. At the strongest imbalance, the median terminal centred errors are $0.09631/0.2456$ for local/global clocks at $\alpha = 0.60$, $0.06324/0.9502$ at $\alpha = 0.75$, and $0.3885/2.228$ at $\alpha = 1.00$. Thus the local-clock advantage in this regime is present for each tested exponent, although its magnitude depends on the decay rate. The larger error of local $\alpha = 1$ relative to local $\alpha = 0.75$ also shows that using participant-specific counts does not eliminate finite-horizon sensitivity to gain decay.

The distinction is not solely an artefact of the unweighted rating-vector norm. In the strongest regime at $\alpha = 0.75$, the median q-weighted population excess log-loss at 10^7 comparisons is 4.220×10^{-6} for the local clock and 2.864×10^{-4} for the global clock. Because this loss weights pairwise predictions by their actual occurrence probabilities, the gap persists even when rarely sampled comparisons contribute less to the aggregate metric. These results concern the displayed finite horizons and matched schedules; they neither establish nonconvergence of the global-clock baseline nor imply that local clocks yield lower error in every exposure regime.

Table 1: Summary of the four pairwise-evaluation streams after preprocessing.

Dataset	Comparisons	Identities	Period	Draws	Identity dynamics
Arena Text	119,870	53	Apr–Jul 2025	Yes	Version-specific
Arena Vision	23,941	17	Feb–Sep 2024	Yes	Version-specific
BASIL	3,551,464	171	Nov 2018–Aug 2026	Rare	Mutable identity
TCEC	29,739	1,595	Jun 2010–Apr 2026	Yes	Version-specific

E.5 Prequential Datasets

Arena Text. Arena Text (LMArena AI, 2025a) consists of human pairwise preferences between responses from frontier language models to common prompts. Outcomes include a preference for either model, a draw, and "both bad". We remove "both bad" observations rather than treating them as draws, leaving 119,870 comparisons among 53 model identities. The data used here span April–July 2025. We do not further subdivide the stream by language or task category.

Arena Vision. Arena Vision (LMArena AI, 2025b) follows the same basic comparison protocol, but prompts may contain both image and textual content. After applying the same preprocessing policy, the dataset contains 23,941 comparisons among 17 model identities, spanning February–September 2024.

BASIL. BASIL (BASIL Ladder, 2026) is a continuous competition between *StarCraft: Brood War* agents. The dataset contains 3,551,464 matches among 171 agent identities from November 2018 through August 2026. Unlike the Arena datasets, an agent can be modified while retaining the same identity. Consequently, a single identity may correspond to an evolving policy and stationary latent strength is less plausible. Draws are possible but rare.

TCEC. The Top Chess Engine Championship (TCEC, 2026) combines league-style all-play-all competition with elimination-style events. We consider traditional chess games from June 2010 through April 2026, yielding 29,739 comparisons among 1,595 engine identities. Engine versions are represented as distinct identities, and successive versions generally do not directly compete, producing a markedly different temporal comparison graph from the Arena datasets. Bradley–Terry was again omitted because fitting the batch model continually over a large participant pool is computationally prohibitive. For one fit, the worst-case time complexity of the chosen implementation (the *MM algorithm*) is linear in the number of comparisons and quadratic in the number of participants. Refitting at short intervals adds an additional scaling factor, hence the difficulty.

E.6 Prequential experiments protocol

Experimental setup For each dataset, observations were converted into temporally ordered tuples of the form

$$(t, i, j, y),$$

where t is the observation time, i and j are the identities being compared, and $y \in [0, 1]$ is the observed outcome from the perspective of i. The resulting sequence was sorted chronologically and partitioned into contiguous training, validation, and test periods using a 60–20–20 split.

The training period is used to initialise the population state of each rating system. In contrast to conventional supervised learning, the online methods continue to update their ratings throughout both validation and test periods. All reported losses are nevertheless strictly prequential: immediately before comparison n, the current ratings are used to form the predicted outcome probability; this prediction is scored against the realised outcome; and only then is the rating state updated. Thus no future outcome is used in forming its own prediction.

Hyperparameters are selected by minimising mean BCE over the validation period. Once selected, the corresponding model is evaluated on the complete temporally ordered stream and performance is reported only over observations belonging to the held-out test period. We report both BCE,

$$-y \log \widehat{p} - (1 - y) \log(1 - \widehat{p}),$$

and the Brier score (an MSE objective of the following form),

$$(y - \widehat{p})^2.$$

Cold-start handling Newly introduced participants create a transient period in which any rating system has little information from which to form a meaningful prediction. Matches involving a participant with fewer than ten previous observations are therefore excluded from the reported loss metrics. These matches are not discarded from the online stream: they still update the internal state of both participants and therefore affect all subsequent predictions. The same rule is applied to every evaluated method and dataset.

TCEC is the regime in which this common warm-up policy deserves the most caution. After preprocessing it contains 29,739 comparisons among 1,595 version-specific engine identities, or about 37 participant appearances per identity on average, and successive versions often occupy different parts of the temporal comparison graph. The ten-observation threshold can therefore consume a larger share of a short-lived identity's history than in the denser datasets. We retain the common rule for comparability across datasets, but treat TCEC as a sparse stress test whose predictive estimates may be more sensitive to cold-start handling and sampling variation.

Hyperparameter optimisation Hyperparameters were optimised independently for each method and dataset using Optuna's multivariate Tree-structured Parzen Estimator (TPE) (Akiba et al., 2019; Bergstra et al., 2011; Falkner et al., 2018), with validation BCE as the optimisation objective. Each optimisation run was allocated 5,000 trials. Table 2 reports the exact final search domains. Parameters spanning multiple scales were sampled log-uniformly where indicated. For CElo-based methods, $\alpha \in [0.5001, 1]$ keeps the search inside the provably convergent regime. For the CUSUM variants, $\delta \in [0, 1]$ reflects the natural residual scale and $\rho \in [0.001, 1]$ spans from an approximately hard effective-count reset to no count reduction. TrueSkill's initial σ was fixed to 1.0 to remove scale redundancy, with the remaining parameters tuned as ratios relative to σ.

We began with broad initial search domains and expanded them when validation-selected hyperparameters reached a search boundary not imposed by theoretical constraints. In such cases, the affected boundary was extended by an order of magnitude, and the optimisation was rerun for that method–dataset pair. Unaffected configurations were not rerun. Table 2 reports the final search domains following these expansions. All decisions to expand the search domains were based exclusively on validation performance.

To ease numerical optimisation, we sometimes use equivalent reparameterisations of the algorithms presented in the main text and Appendix D. In particular, the experimental CElo gain uses shared parameters $a_i = M$ and $c_i = c$ and can be written as

$$\lambda_{n,i} = \frac{c}{(M + \nu_i(n))^\alpha} = \eta_0 \left(\frac{M}{M + \nu_i(n)} \right)^\alpha,$$

where $\eta_0 = c/M^\alpha$ is the initial gain. For CElo, we optimise $(\eta_0, \alpha, T_{1/2})$, where

$$T_{1/2} = M(2^{1/\alpha} - 1)$$

is the number of participant appearances required to halve the initial gain. Thus, $M = T_{1/2}/(2^{1/\alpha} - 1)$ and $c = \eta_0 M^\alpha$. For both C-CElo variants, we instead optimise (η_0, α, M) directly, again setting $c = \eta_0 M^\alpha$. Their gains use the resettable effective count κ_i in place of the lifetime participation count ν_i, with the respective reset rules defined in Appendix D.1. These parameter conversions leave the corresponding gain formulas unchanged.

Table 2: Hyperparameter descriptions and final Optuna search domains. Dataset-specific ranges are shown explicitly. "log" denotes log-uniform sampling; all other continuous domains are sampled linearly.

Parameter	Description	Search domain
Elo		
η	Learning rate (fixed step size)	$[10^{-4}, 5.0]$ (log)
CElo		
η_0	Initial learning rate	$[10^{-3}, 30.0]$ (log)
α	Polynomial decay rate	$[0.5001, 1.0]$
half-life	Participant appearances until initial gain halves	$[0.1, 10000.0]$ (log)
C-CElo variants		
η_0	Initial learning rate	$[10^{-2}, 10.0]$ (log)
α	Polynomial decay rate	$[0.5001, 1.0]$
M	Pre-observation clock offset	$[0.1, 1000.0]$ (log)
δ	CUSUM drift tolerance	$[0, 1]$
h (C-CElo)	CUSUM alarm threshold	Arena Text/BASIL: $[0.1, 100]$ (log) Other datasets: $[0.1, 10]$ (log)
h (zero anchor)	CUSUM alarm threshold	Arena Text: $[0.1, 1000]$ (log) Other datasets: $[0.1, 10]$ (log)
ρ	Effective-count reset factor	$[0.001, 1.0]$
Glicko-1		
initial ϕ	Initial rating deviation	$[10^{-2}, 10.0]$ (log)
max. ϕ ratio	Cap on deviation relative to initial ϕ	TCEC: $[1, 100]$ (log) Other datasets: $[1, 10]$ (log)
decay	Rating-deviation growth factor	$[10^{-4}, 50.0]$ (log)
days / period	Window length for batched updates	$\{0, \ldots, 365\}$ (integer)
Glicko-2		
initial ϕ	Initial rating deviation	BASIL: $[10^{-2}, 100]$ (log) Other datasets: $[10^{-2}, 10]$ (log)
initial vol.	Initial rating volatility	$[10^{-4}, 2.0]$ (log)
τ	Volatility constraint parameter	TCEC: $[10^{-4}, 50]$ (log) Other datasets: $[10^{-3}, 5]$ (log)
days / period	Window length for batched updates	$\{0, \ldots, 365\}$ (integer)
TrueSkill		
β/σ	Dynamics variance relative to scale	Arena Text/Vision: $[0.05, 50]$ (log) BASIL/TCEC: $[0.05, 5]$ (log)
ν/σ	Update-step variance relative to scale	$[10^{-4}, 1.0]$ (log)
draw margin/σ	Threshold for tie probability	Arena Text/Vision: $[0, 20]$ BASIL/TCEC: $[0, 2]$
Bradley–Terry		
reg. wins	Prior pseudo-observations (regularisation)	$[10^{-3}, 10^4]$ (log)
days / period	Re-fitting window interval	$\{1, 7, 30, 182, 365\}$ (categorical)

Table 3 reports the final configuration selected on each validation set. No test-set observations were used for hyperparameter selection.

Table 3: Validation-selected hyperparameters. Each block reports the selected configuration for one rating method. A dash denotes a method–dataset combination without an extracted value.

Parameter	Arena Text	Arena Vision	BASIL	TCEC
Elo				
η	0.0066827	0.0065879	0.011727	0.26415
CElo				
η_0	0.030469	0.011837	2.1207	2.9619
α	0.58054	0.99976	0.50023	0.94508
half-life	161.5	928.1	2.0487	1.4138
C-CElo (zero anchor)				
η_0	0.028351	0.54906	1.3392	2.8366
α	0.58503	0.99965	0.62382	0.93118
M	84.755	14.721	1.5863	1.2786
δ	0.81289	0.72711	0.098394	0.59765
h	0.9205	0.17781	9.9382	0.25962
ρ	0.21501	0.061515	0.0040598	0.018667
C-CElo				
η_0	0.036087	0.090753	1.2498	3.1104
α	0.63964	0.98083	0.65901	0.98534
M	65.137	1.5302	2.8166	1.4181
δ	0.068553	0.46186	0.10014	0.72511
h	10.163	0.31006	10.006	0.13339
ρ	0.011901	0.0014942	0.0017641	0.0011597
Glicko-1				
initial ϕ	0.36588	0.36348	3.8925	2.8268
max. ϕ ratio	6.5954	1.6916	2.9815	60.522
decay	0.05181	0.053273	0.019677	0.10352
days / period	13	2	1	1
Glicko-2				
initial ϕ	0.42547	0.30751	12.399	2.9919
days / period	13	2	1	1
initial vol.	0.065053	0.051836	0.016453	0.091036
τ	0.012333	0.096934	0.49123	0.0011785
TrueSkill				
β/σ	6.3516	11.054	1.0497	0.52098
ν/σ	0.0068047	0.014254	0.0036009	0.012876
draw margin / σ	2.3109	10.396	0.000263	0.11964
Bradley–Terry				
days / period	1	1	1	–
reg. wins	34.775	94.119	0.33904	–

Paired bootstrap uncertainty Because observations are temporally ordered, uncertainty in the difference between each method and Elo was estimated using a paired moving-block bootstrap rather than by resampling individual comparisons independently, following (Künsch, 1989). For each test stream, contiguous blocks of 500 observations were resampled to construct 20,000 bootstrap replicates. Within each replicate, the mean BCE of a given method and the Elo baseline were evaluated on the same resampled observations, and their difference

$$\Delta_{\mathrm{BCE}} := \mathrm{BCE}_{\mathrm{method}} - \mathrm{BCE}_{\mathrm{Elo}}$$

was recorded. Negative values therefore favour the method under comparison. The reported 95% intervals are the corresponding empirical bootstrap intervals. The rating systems are not re-

fit within bootstrap replicates, so the intervals quantify uncertainty in held-out loss conditional on the validation-selected configuration. Because the held-out streams may be nonstationary, we do not establish the dependence conditions required for nominal bootstrap coverage; the intervals are therefore interpreted descriptively/approximately.

Because a fixed block length corresponds to very different fractions of the four held-out streams, we additionally audited block-length sensitivity using blocks equal to approximately 0.5%, 1%, and 2% of each test stream, alongside the 500-comparison specification, with 10,000 replicates per setting. The CI's qualitative relation to zero was unchanged across these block lengths for all method–dataset comparisons except Bradley–Terry on Arena Text and Arena Vision, where the interval alternated between lying below zero and overlapping zero. This sensitivity does not affect the CElo-versus-Elo conclusions reported in Section 4.3.

F Complete Prequential Results

Figure 7 shows the applied test-set BCE differences relative to Elo. Negative values favour the method under comparison; intervals are paired moving-block 95% bootstrap intervals. Table 4 reports the corresponding exact BCE and Brier scores. The CUSUM variants are intended only as engineering wrappers around the convergent CElo core: they address diminishing-gain stagnation by resetting effective local counts while leaving the rating update unchanged. Across the four streams, the average within-dataset BCE ranks are 2.50 for C-CElo (zero anchor) and 2.75 for C-CElo. We treat these as exploratory nonstationary stress-test results rather than part of the convergence contribution.

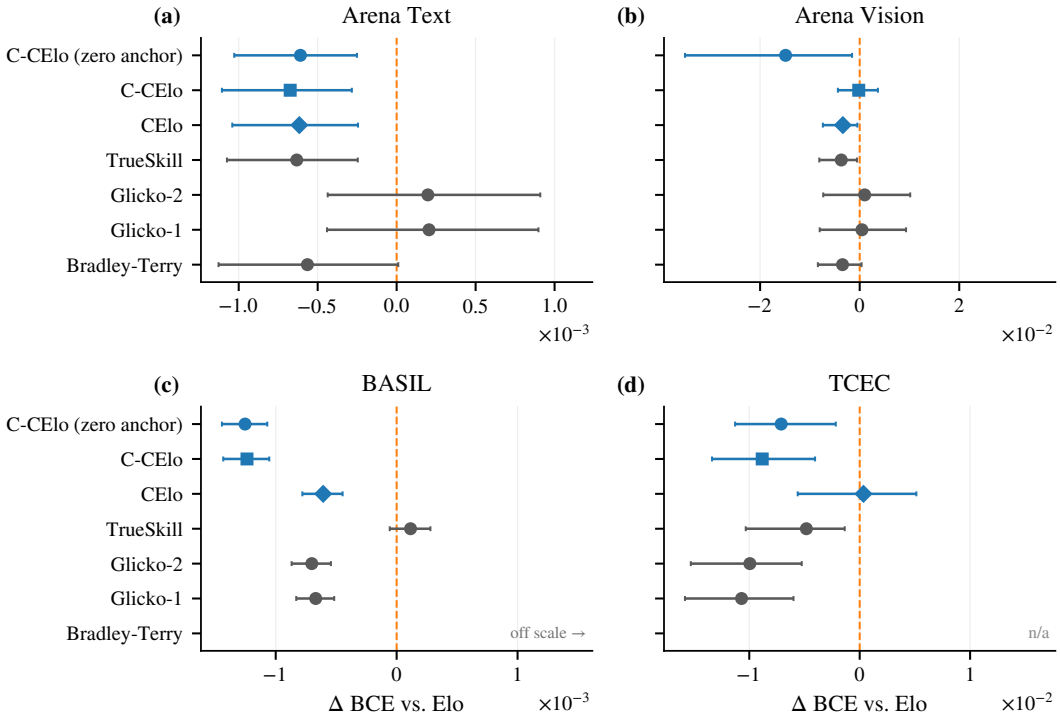

Figure 7: Held-out predictive performance on four temporally ordered evaluation streams. Points show paired differences in test BCE relative to Elo and bars show paired moving-block 95% bootstrap intervals (500-comparison blocks; 20,000 bootstrap replicates). Negative values indicate lower BCE than Elo; the dashed vertical line is the Elo baseline at zero. Bradley–Terry results outside a panel's display range are marked as off-scale, with exact values reported in Table 4.

Table 4: Complete held-out predictive results. Best BCE on each dataset is bold. ΔBCE is relative to Elo with paired moving-block 95% bootstrap intervals (500-comparison blocks; 20,000 bootstrap replicates).

Method	Brier	BCE	Δ BCE vs. Elo [95% CI]
Arena Text			
Elo	0.188049	0.658624	+0.000000 [+0.000000, +0.000000]
CElo	0.187753	0.658008	-0.000616 [-0.001041, -0.000245]
C-CElo (zero anchor)	0.187757	0.658014	-0.000609 [-0.001028, -0.000251]
C-CElo	0.187723	**0.657950**	-0.000674 [-0.001106, -0.000284]
Glicko-1	0.188147	0.658829	+0.000205 [-0.000441, +0.000898]
Glicko-2	0.188144	0.658821	+0.000198 [-0.000437, +0.000910]
TrueSkill	0.187744	0.657991	-0.000632 [-0.001075, -0.000246]
Bradley–Terry	0.187778	0.658059	-0.000565 [-0.001128, +0.000010]
Arena Vision			
Elo	0.166676	0.614772	+0.000000 [+0.000000, +0.000000]
CElo	0.165184	0.611394	-0.003378 [-0.007396, -0.000468]
C-CElo (zero anchor)	0.160467	**0.599890**	-0.014882 [-0.035094, -0.001532]
C-CElo	0.166498	0.614582	-0.000190 [-0.004377, +0.003659]
Glicko-1	0.167190	0.615223	+0.000451 [-0.008034, +0.009310]
Glicko-2	0.167429	0.615778	+0.001006 [-0.007324, +0.010148]
TrueSkill	0.165086	0.611072	-0.003699 [-0.008117, -0.000541]
Bradley–Terry	0.165218	0.611317	-0.003454 [-0.008399, +0.000372]
BASIL			
Elo	0.194131	0.568374	+0.000000 [+0.000000, +0.000000]
CElo	0.193875	0.567766	-0.000608 [-0.000780, -0.000447]
C-CElo (zero anchor)	0.193606	**0.567120**	-0.001253 [-0.001445, -0.001070]
C-CElo	0.193614	0.567136	-0.001238 [-0.001434, -0.001053]
Glicko-1	0.193838	0.567704	-0.000669 [-0.000830, -0.000516]
Glicko-2	0.193827	0.567672	-0.000702 [-0.000868, -0.000544]
TrueSkill	0.194029	0.568489	+0.000115 [-0.000057, +0.000279]
Bradley–Terry	0.200725	0.583448	+0.015074 [+0.014549, +0.015590]
TCEC			
Elo	0.105142	0.662705	+0.000000 [+0.000000, +0.000000]
CElo	0.104201	0.663046	+0.000341 [-0.005614, +0.005132]
C-CElo (zero anchor)	0.101647	0.655595	-0.007111 [-0.011282, -0.002166]
C-CElo	0.100838	0.653877	-0.008828 [-0.013374, -0.004042]
Glicko-1	0.099869	**0.652007**	-0.010698 [-0.015824, -0.005994]
Glicko-2	0.100089	0.652759	-0.009946 [-0.015292, -0.005248]
TrueSkill	0.102528	0.657877	-0.004828 [-0.010320, -0.001353]
Bradley–Terry	–	–	–