PARETO CALIBRATION FOR PAIRWISE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

When learning a score from pairwise comparisons by minimizing a surrogate margin loss, one naturally expects the learned ranking to respect the data: if all comparisons, or even only a strict majority of them, prefer a to b, then the score should prefer a as well, known in classical social choice theory as Pareto optimality (PO) and pairwise majority consistency (PMC). Somewhat unexpectedly, recent results show that, unless the comparisons are spherically symmetric, no reasonable convex loss can enforce either axiom, even in the simple setting of linear social choice (Ge et al., 2024; Hollender & Kraiczy, 2026). We reexamine these negative and positive results and identify the missing link as classification calibration. Spherical symmetry is special precisely because it decomposes the pairwise objective into a mixture of such classification problems, so calibration of the loss transfers to PO and PMC. In contrast, without posing restrictions on data, we show no condition on the surrogate alone can restore PO. Then, we introduce Pareto calibration, a joint condition on the loss and the comparison measure that is necessary and sufficient for PO. We give a sufficient condition that can be checked from the observed pairs and the slope of the loss, and show that any surrogate that is calibrated for the classification task can be made to satisfy it on every distribution, either by truncating the loss or by reweighting the data. Based on the theoretical insight build off from this condition, we introduce loss truncation and data reweighing, which improves empirical Pareto calibration result across different model sizes and learning objectives.

1 INTRODUCTION

Many learning problems can be unified as learning from pairwise data, including preference optimization, ranking, and both binary and multiclass classification. In pairwise data, each observation records that an alternative a is preferred to an alternative b, and we aim to learn a score function h that ranks alternatives by their scores. One way to learn h is to minimize the 0-1 loss, which charges a mistake whenever h ranks b above a. However, the 0-1 loss is discontinuous and hard to optimize. Therefore, we often minimize a (convex) surrogate loss ϕ on the difference $h(a) - h(b)$. By pairing each example against a dummy alternative, classification is included in this framework. By parameterizing the score with language-model logits, reward modeling and preference optimization are also included (Christiano et al., 2017; Rafailov et al., 2023).

In some of these settings, the role of the surrogate ϕ is well understood. For classification, the goal is to achieve Bayes consistency, and calibration characterizes when minimizing a tractable ϕ recovers the Bayes-optimal decision (Bartlett et al., 2006; Tewari & Bartlett, 2007). The same cannot be said once the surrogate is used to aggregate preferences. Recent work on preference-based optimization has found that minimizing ϕ can violate some very basic and intuitive properties, such as Pareto Optimality (PO) and Pairwise Majority Consistency (PMC) (Arrow et al., 2010; Brandt et al., 2016; Ge et al., 2024; Hollender & Kraiczy, 2026).

Pareto Optimality (PO) asks only that the learnt score h respects the unanimous preference, i.e. if every observed comparison between a and b prefers a, the learned score should satisfy $h(a) > h(b)$. A stronger guarantee is Pairwise Majority Consistency (PMC), which asks that the learnt score h respects the pairwise majority preference, i.e. if every strict pairwise majority between a and b prefers a, the learned score should satisfy $h(a) > h(b)$. PO is automatically satisfied when PMC is satisfied. Prior work gives positive guarantees when comparisons are distributed on the unit sphere,

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.

and shows that a fixed loss cannot guarantee PO for every comparison distribution (Ge et al., 2024; Hollender & Kraiczy, 2026).

Before asking when a surrogate can satisfy an intuitive requirement such as PO, we first ask what it means for ϕ to stand in for the 0-1 comparison loss. We first show that minimizing that 0-1 loss is equivalent to the generalized Kemeny rule (Kemeny, 1959). When the pairwise data are induced by a complete tournament, i.e. every pair of alternatives is compared, it is known that the Kemeny rule satisfies both PO and PMC (Brandt et al., 2016). Thus, in learning from pairwise data, minimizing a surrogate ϕ can be viewed as a convex relaxation of the Kemeny rule.

We then revisit the positive and negative results of Ge et al. (2024) and Hollender & Kraiczy (2026). What makes their positive result work is a link to calibration in classification, which we call classification calibration for the rest of the paper. Under spherically symmetric comparisons the pairwise objective splits into a mixture of binary conditional risks, and PMC (hence PO) holds as soon as ϕ is classification-calibrated. However, the spherically symmetric data condition turns out to be crucial. Indeed, we show that no property of the loss by itself can guarantee even Pareto optimality. We therefore introduce Pareto calibration (PC), a joint condition on the loss and the data distribution, and show that it is necessary and sufficient for PO.

Unfortunately, PC is not easy to verify, because it asks us to compare two infima over all scores. We thus relax it to a sufficient condition that uses only the observed pairwise data and the slope of ϕ. Any classification-calibrated loss can be made to satisfy this condition on any data distribution, either by truncating the loss or by reweighting the data. The two methods are equivalent through a duality, and this also justifies existing data-reweighting heuristics in preference optimization.

2 PRELIMINARY

Fix a finite set of alternatives $\mathcal{X}$. Let ρ be finite and nonnegative on $\mathcal{X}\times\mathcal{X}$, where $\rho_{ab} := \rho(\{(a,b)\})$ is the proportion of comparisons in which a is preferred to b. Both ρ_{ab} and ρ_{ba} may be positive, and we set $\rho_{aa} = 0$ throughout. We call a pair (a,b) oriented if $\rho_{ab} > 0$, with notation $a \to b$.

A score is a function $h : \mathcal{X} \to \mathbb{R}$. For an oriented pair (a,b), define the margin $m_{ab}(h) := h(a) - h(b)$. Positive margin means that h agrees with the orientation $a \to b$. Given a decreasing margin loss $\phi : \mathbb{R} \to \mathbb{R}$, the aggregate pairwise surrogate risk is

$$L_\rho^\phi(h) := \sum_{a,b\in\mathcal{X}} \rho_{ab}\,\phi\big(m_{ab}(h)\big). \tag{1}$$

Define the net preference $\Delta_\rho(a,b) := \rho_{ab} - \rho_{ba}$ and $\rho^+_{ab} := \rho_{ab}\mathbf{1}_{\{\Delta_\rho(a,b)>0\}}$, the weight carried by the strict majority orientation. When $\Delta_\rho(a,b) > 0$, the population places more weight on $a \to b$ than on $b \to a$.

The two definitions below formalize how faithfully an aggregate score must reproduce the population's pairwise preferences. The Pairwise majority consistency (PMC) is stronger, asking the score to agree with every strict-majority preference, i.e. whenever more comparisons orient a over b than the reverse. Pareto optimality (PO) is weaker, constraining the score only on pairs where the preference is unanimous, so that no comparison points the other way. Since unanimity implies a strict majority, PMC implies PO.

Definition 2.1 (Pairwise Majority Consistency (PMC))**.** *A score h satisfies pairwise majority consistency (PMC) for ρ if $\rho^+_{ab} > 0$ implies $h(a) > h(b)$ for every pair $a, b \in \mathcal{X}$.*

Definition 2.2 (Pareto Optimality (PO))**.** *A score h is Pareto optimal (PO) for ρ if $\Delta_\rho(a,b) = \rho^+_{ab} > 0$ implies $h(a) > h(b)$ for every pair $a, b \in \mathcal{X}$.*

The condition $\Delta_\rho(a,b) = \rho^+_{ab} > 0$ says $\rho_{ab} > 0 = \rho_{ba}$, i.e. every comparison between a and b prefers a. We call such a pair (a,b) unanimous and write $E_U := \{(a,b) \in \mathcal{X}^2 : \Delta_\rho(a,b) = \rho^+_{ab} > 0\}$, $E_N := \mathrm{supp}(\rho) \setminus E_U$. as the set of unanimous and non-unanimous pairs. For $e \in \mathrm{supp}(\rho)$ we abbreviate $\rho_e := \rho(e)$.

Whether a single score can agree with all unanimous preferences at once is dependent on how these preferences interact. It is therefore natural to restrict the discussion to the case where the mildest condition, Pareto Optimality, is realizable.

Definition 2.3 (Pareto Realizable). *Let G_U be a directed graph with an arrow $a \to b$ for every $(a, b) \in E_U$. ρ is Pareto realizable whenever G_U has no directed cycle.*

By instantiating the set of alternatives, we recover several important learning problems, including binary and multi-class classification, ranking, linear social choice (Ge et al., 2024), and preference-based alignment of language models. We detail binary classification and two alignment objectives here, and defer multi-class classification, ranking, and linear social choice to Appendix A.

Example 2.1 (Binary Classification). *Let $\mathcal{X}$ be a feature space and let P be a distribution on $\mathcal{X} \times \{-1, +1\}$. The alternative set is $\mathcal{A}_{\text{bin}} := \mathcal{X} \times \{-1, +1\}$. Define the map $T_{\text{bin}} : \mathcal{A}_{\text{bin}} \to \mathcal{A}^2_{\text{bin}}$ by $T_{\text{bin}}(x, y) = \big((x, y), (x, -y)\big)$, where $x \in \mathcal{X}$, $y \in \{-1, +1\}$. Let $\rho_{\text{bin}} := P \circ T^{-1}_{\text{bin}}$ be the induced comparison measure. For a score $h : \mathcal{A}_{\text{bin}} \to \mathbb{R}$, write $s_y(x) := h(x, y)$ and $f(x) := s_{+1}(x) - s_{-1}(x)$. Then*

$$L^{\phi}_{\rho_{\text{bin}}}(h) = \mathbb{E}_{(X,Y)\sim P}\big[\phi(m_{T_{\text{bin}}(X,Y)}(h))\big] = \mathbb{E}\big[\phi(s_Y(X) - s_{-Y}(X))\big] = \mathbb{E}\big[\phi(Y f(X))\big].$$

Thus, the pairwise risk reduces exactly to the binary margin risk $\mathbb{E}[\phi(Y f(X))]$ used in binary classification (Bartlett et al., 2006).

To interpret PMC and PO in the binary setting, write $\eta(x) := P(Y = +1 \mid X = x)$ for each x with $P(X = x) > 0$, so that $\Delta_{\rho_{\text{bin}}}\big((x, +1), (x, -1)\big) = P(X = x)\big(2\eta(x) - 1\big)$. PMC at x then requires the more probable label to receive the higher score: $\eta(x) > \frac{1}{2} \implies s_{+1}(x) > s_{-1}(x)$ and $\eta(x) < \frac{1}{2} \implies s_{-1}(x) > s_{+1}(x)$, which is exactly the Bayes sign rule $\text{sign}(2\eta(x) - 1)$. PO is the restriction of this rule to deterministic labels: $\eta(x) \in \{0, 1\}$.

The same template also captures preference-based alignment of language models, where comparisons are human judgments between candidate responses.

Example 2.2 (Reward learning with the Bradley-Terry model). *Let $\mathcal{X}$ be a space of prompts and $\mathcal{Y}$ a space of responses, and take the alternative set $\mathcal{A}_{\text{BT}} := \mathcal{X} \times \mathcal{Y}$ of prompt-response pairs. Preference data is a distribution P over triples (X, Y_w, Y_l), where for prompt X the response Y_w is preferred to Y_l. Define the map $T_{\text{BT}}(x, y_w, y_l) := \big((x, y_w), (x, y_l)\big)$ sending each preference to its oriented pair, and let $\rho_{\text{BT}} := P \circ T^{-1}_{\text{BT}}$ be the induced comparison measure. A score $h : \mathcal{A}_{\text{BT}} \to \mathbb{R}$ is a reward model $r(x, y) := h(x, y)$, with margin $m_{(x,y_w),(x,y_l)}(h) = r(x, y_w) - r(x, y_l)$. The Bradley-Terry model sets $\mathbb{P}(y_w \succ y_l \mid x) = \big(1 + e^{-(r(x,y_w)-r(x,y_l))}\big)^{-1}$, and reward learning minimizes its negative log-likelihood. With the logistic loss $\phi(m) = \log(1 + e^{-m})$,*

$$L^{\phi}_{\rho_{\text{BT}}}(h) = \mathbb{E}_{(X,Y_w,Y_l)\sim P}\big[\phi\big(r(X, Y_w) - r(X, Y_l)\big)\big] = -\mathbb{E}\big[\log \mathbb{P}(Y_w \succ Y_l \mid X)\big],$$

which is exactly the Bradley-Terry reward-modeling objective (Christiano et al., 2017).

Likewise, by parameterizing the score h with the model itself, i.e. $h_\theta(x, y) := \beta \log \frac{\pi_\theta(y|x)}{\pi_{\text{ref}}(y|x)}$, we recover the direct preference optimization objective (Rafailov et al., 2023).

2.1 The $0-1$ Loss and the Kemeny Rule

Since we phrase aggregation as minimizing a risk L^{ϕ}_{ρ} (Equation 1), a natural starting point is the 0-1 loss, which charges a unit penalty for every comparison the score gets wrong. Under this loss the framework recovers a classical social choice rule: the Kemeny rule (Kemeny, 1959).

Consider $\phi_{0\text{-}1}(\xi) = \mathbf{1}_{\xi \le 0}$, which counts ties as mistakes, so that $L^{0\text{-}1}_{\rho}(h) = \sum_{a,b} \rho_{ab} \mathbf{1}_{h(a) \le h(b)}$. A strict total order σ on $\mathcal{X}$ is a ranking that places every pair $a \neq b$ in exactly one of $a \succ_\sigma b$ or $b \succ_\sigma a$. Its discordance is $K_\rho(\sigma) := \sum_{a \neq b} \rho_{ab} \mathbf{1}_{b \succ_\sigma a}$, the total ρ-mass of comparisons $a \to b$ that σ reverses.

Definition 2.4 (Generalized Kemeny rule). *Let $\Sigma(\mathcal{X})$ denote the set of all strict total orders on $\mathcal{X}$. The generalized Kemeny rule returns any discordance-minimizing order, $\sigma^\star \in \text{argmin}_{\sigma \in \Sigma(\mathcal{X})} K_\rho(\sigma)$ with ties among minimizers broken arbitrarily.*

Proposition 2.1 (0-1 risk equals generalized Kemeny). *For every ρ and every score h, and every strict total order σ that linearly extends the strict relation induced by h, $K_\rho(\sigma) \le L^{0\text{-}1}_{\rho}(h)$ with equality when h is injective. Consequently, $\inf_{h:\mathcal{X}\to\mathbb{R}} L^{0\text{-}1}_{\rho}(h) = \min_\sigma K_\rho(\sigma)$ where the minimum on the right ranges over strict total orders on $\mathcal{X}$.*

Therefore, minimizing the 0-1 loss is equivalent to computing a Kemeny order. When ρ is induced by complete rankings over $\mathcal{X}$, the Kemeny rule is known to satisfy both PMC and PO (Brandt et al., 2016). For incomplete rankings, the same holds under a mild monotonicity condition on ρ.

Definition 2.5 (Monotonicity)**.** *Let $a, b \in \mathcal{X}$ form a strict majority pair, i.e. $\Delta_\rho(a,b) > 0$. Then ρ is monotone at (a,b) if $\rho_{db} \geq \rho_{da}$ for every $d \in \mathcal{X} \setminus \{a,b\}$, and $\rho_{ae} \geq \rho_{be}$ for every $e \in \mathcal{X} \setminus \{a,b\}$.*

Monotonicity at every majority pair rules out Condorcet cycles, the classic social choice dilemma in which majorities prefer a over b, b over c, and c over a simultaneously. It forces the strict-majority relation $a \succ_M b \Leftrightarrow \Delta_\rho(a,b) > 0$ to be transitive, so a consistent ranking always exists.

Theorem 2.1. *Fix a pair (a,b) such that $\Delta_\rho(a,b) > 0$. If ρ is monotone at (a,b), then every Kemeny minimizer $\sigma^\star$ ranks $a \succ_{\sigma^\star} b$, and every minimizer $h^\star$ of $L_\rho^{0\text{-}1}$ over all scores satisfies $h^\star(a) > h^\star(b)$. Consequently, 1) if ρ is monotone at every strict majority pair, then every Kemeny minimizer and every 0-1 minimizer satisfies pairwise majority consistency. 2) if ρ is monotone at every unanimous pair (every (a,b) with $\rho_{ba} = 0 < \rho_{ab}$), then every such minimizer is Pareto optimal.*

Minimizing the 0-1 loss, equivalently computing a Kemeny order, is NP-hard (Bartholdi III et al., 1989). In learning settings, one typically replaces $\phi_{0\text{-}1}$ by a convex surrogate margin loss ϕ and minimizes L_ρ^ϕ instead. Through the equivalence above, this makes surrogate risk minimization a convex relaxation of the Kemeny rule. The relaxation, however, need not preserve the axioms that make the exact rule desirable.

We next ask whether the monotonicity guarantee extends beyond the 0-1 loss to a general convex surrogate ϕ. Write $\mathbb{R}_{--} := (-\infty, 0)$ and $\mathbb{R}_{++} := (0, \infty)$, and say that ϕ satisfies $\phi(\mathbb{R}_{--}) > \phi(\mathbb{R}_{++})$ if $\phi(u) > \phi(v)$ for every $u < 0 < v$, i.e. every negative margin incurs strictly more loss than every positive one.

Theorem 2.2. *Fix a pair (a,b) with $\Delta_\rho(a,b) > 0$ and suppose ρ is monotone at (a,b). Let $\phi : \mathbb{R} \to \mathbb{R}$ be nonincreasing with $\phi(\mathbb{R}_{--}) > \phi(\mathbb{R}_{++})$, and suppose L_ρ^ϕ attains its minimum over all scores. Then every minimizer $h^\star$ satisfies $h^\star(a) \geq h^\star(b)$. If, in addition, ϕ is differentiable on $\mathbb{R}$ with $\phi'(0) < 0$ then $h^\star(a) > h^\star(b)$.*

3 When can a Loss Enforce PMC and PO?

The monotonicity results of the previous section required conditions on both ρ and ϕ. To understand what is possible in general, we turn to a structured instance, linear social choice, where parts of the answer are already known. We first extend the negative results to all comparison measures, showing that no single loss guarantees PO universally. We then turn to the positive results in linear social choice and identify a crucial connection with binary classification calibration, which motivates the general condition developed in Section 4.

Linear Social Choice Linear social choice (Ge et al., 2024; Hollender & Kraiczy, 2026) is a structured instance of Equation 1 in which the alternative set is $\mathcal{X} = \mathbb{S}^{d-1}$, the unit sphere in $\mathbb{R}^d$. Each voter i has a preferred direction $v_i \in \mathbb{S}^{d-1}$ and prefers x over $-x$ for every x in its open halfspace $H_i := \{x \in \mathbb{S}^{d-1} : \langle v_i, x\rangle > 0\}$, contributing the oriented comparison $(x, -x)$. The comparison measure is $\rho(A) = \sum_{i=1}^n \sigma\{x \in H_i : (x,-x) \in A\}$ for every measurable $A \subseteq \mathcal{X}^2$, where σ is the normalized uniform measure on $\mathbb{S}^{d-1}$. Parameterizing linear scores as $h_\theta(x) = \frac{1}{2}\langle \theta, x\rangle$, so that their antipodal margin is $\langle \theta, x\rangle$, gives the risk $L_{\text{unif}}^\phi(\theta) := \sum_{i=1}^n \int_{\mathbb{S}^{d-1} \cap \{\langle v_i, x\rangle > 0\}} \phi\big(\langle \theta, x\rangle\big)\, d\sigma(x)$, where the decreasing loss ϕ rewards aligning θ with the directions each voter prefers. This parameterization preserves the same margin loss ϕ as Equation 1, since the class of linear rankings is unchanged by the factor $1/2$. A direction x is unanimous when $\langle v_i, x\rangle > 0$ for every voter i, and PO asks the minimizer $\theta^\star$ to satisfy $\langle \theta^\star, x\rangle > 0$ at every such x.

In this restricted setting, Hollender & Kraiczy (2026) showed that no strictly convex, lower-bounded, and differentiable loss with $\phi'(0) < 0$ can guarantee PO for every choice of voter directions. We first generalize this impossibility from linear scores to the set of all scores $\mathcal{H}_{\text{all}} := \{h : \mathcal{X} \to \mathbb{R}\}$.

Theorem 3.1. *Let ϕ be continuous, convex, and coercive with a unique minimizer. Then there exists a PO-realizable measure ρ such that every minimizer of L_ρ^ϕ over $\mathcal{H}_{\text{all}}$ violates Pareto optimality.*

While we defer the proof to the appendix, the following example illustrates the main idea.

Example 3.1. *Let $\mathcal{X} = \{a, b, c\}$ and set $\rho_{ab} = 2$, $\rho_{bc} = \rho_{ca} = 12$, $\rho_{cb} = \rho_{ac} = \frac{1}{2}$ with zero mass on the remaining pairs. The only unanimous edge is $a \to b$, so ρ is Pareto realizable. The loss $\phi(\xi) = (1-\xi)^2$ is continuous, strictly convex, and coercive. Write $x := h(a) - h(b)$ and $y := h(b) - h(c)$. The risk is $F(x, y) = 2(1-x)^2 + 12(1-y)^2 + \frac{1}{2}(1+y)^2 + 12(1+x+y)^2 + \frac{1}{2}(1-x-y)^2$. Solving this gives the unique solution $x = -38/33$ and $y = 19/33$, which violates PO.*

On the positive side, Hollender & Kraiczy (2026) showed that the uniform structure of ρ over the sphere is enough, in the sense that PO holds for every strictly convex, lower-bounded, and differentiable loss with $\phi'(0) < 0$ when the data is uniform. Coincidentally, this condition on ϕ is precisely the classical condition for a convex surrogate to be classification-calibrated (Bartlett et al., 2006), the property that justifies losses such as logistic or exponential as surrogates for the intractable 0-1 loss in binary classification.

Classification Calibration In the binary classification setup of Example 2.1, for $\eta \in (0, 1)$ write $\mathbb{P}(Y = +1) = \eta$ for the (conditional) probability of a positive label. The (conditional) ψ-risk at margin $\alpha \in \mathbb{R}$ is $C_\eta^\psi(\alpha) := \eta\,\psi(\alpha) + (1-\eta)\,\psi(-\alpha) = \mathbb{E}[\psi(Y\alpha) \mid \mathbb{P}(Y = +1) = \eta]$. Write $H^\psi(\eta) := \inf_{\alpha\in\mathbb{R}} C_\eta^\psi(\alpha)$ and $H^{\psi,-}(\eta) := \inf_{\alpha:\,\alpha(2\eta-1)\le 0} C_\eta^\psi(\alpha)$. Here $H^\psi(\eta)$ is the smallest conditional risk attainable at positive-label probability η, while $H^{\psi,-}(\eta)$ is the smallest attainable when the margin has the wrong sign relative to the Bayes rule $\mathbf{sign}(2\eta-1)$. Calibration requires the wrong-signed margin to be strictly costlier whenever $\eta \neq \frac{1}{2}$, so that any surrogate-optimal margin automatically carries the Bayes-optimal sign.

Definition 3.1 (Classification calibration, Bartlett et al. 2006)**.** *The loss ψ is classification-calibrated if $H^{\psi,-}(\eta) > H^\psi(\eta)$ for every $\eta \neq \frac{1}{2}$.*

For convex margin losses, classification calibration has a simple characterization.

Lemma 3.1 (Bartlett et al. 2006)**.** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex and differentiable at 0. Then ϕ is classification-calibrated if and only if $\phi'(0) < 0$.*

To see how classification calibration connects to the positive result in linear social choice, we observe that the linear social choice risk decomposes into a mixture of such conditional risks.

Lemma 3.2. *For $x \in \mathbb{S}^{d-1}$, let $n_+(x) := |\{i : \langle v_i, x\rangle > 0\}|$ and $\eta(x) := n_+(x)/n$. Then for every $\theta \in \mathbb{R}^d$, $L^\phi_{\mathrm{unif}}(\theta) = \frac{n}{2}\int_{\mathbb{S}^{d-1}} C^\phi_{\eta(x)}\big(\langle\theta, x\rangle\big)\, d\sigma(x)$.*

Through this decomposition, each direction x behaves like an independent binary classification problem with positive-label probability $\eta(x)$. On a unanimous direction every voter agrees, so $\eta(x) = 1$, and a calibrated ϕ forces the pointwise-optimal margin to be strictly positive, which is exactly PO.

With this observation and by classification calibration, we can strenghthen the result of Hollender & Kraiczy (2026) to a PMC guarantee, provided that the scores are not necessarily linear.

Lemma 3.3. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be classification-calibrated, bounded below, and lower semicontinuous, and assume L^ϕ_ρ attains its minimum on $\mathcal{H}_{\mathrm{all}} := \{h : \mathbb{S}^{d-1} \to \mathbb{R}\}$. Then every minimizer $h^\star$ satisfies pairwise majority consistency for σ-almost every $x \in \mathbb{S}^{d-1}$.*

Much of the positive results here rely on the connection between PMC/PO and classification calibration, which is specific to the symmetric structure of the sphere. We give a more in-depth discussion of this connection in the appendix. For example, multi-class classification induces a monotone ρ_{mc}, so PMC and PO are achievable. But PMC is strictly stronger than the existing multi-class calibration of Tewari & Bartlett (2007).

We also extend further the uniform-data results of Hollender & Kraiczy (2026) in two directions, (i) the voters may be drawn from an arbitrary distribution ν on the unit sphere, and (ii) the comparison directions may follow any spherically symmetric law μ on $\mathbb{R}^d$ (e.g. uniform on the sphere or a standard Gaussian). This sharpens the special case of Hollender & Kraiczy (2026). We defer the details to Appendix E.1.

4 Pareto Calibration

The negative and positive results of the previous section motivate working with a calibration-style condition. A general comparison measure ρ, however, does not admit the mixture-of-binary-

classification decomposition that made classification calibration apply, so we introduce a new definition of calibration for general comparison measures. We then show that this is a necessary and sufficient condition for PO, the milder of the two guarantees.

Definition 4.1. *For an oriented pair $e = (a, b)$, write $\mathcal{C}_e := \{h \in \mathcal{H}_{\text{all}} : m_e(h) \leq 0\}$. The pair (ρ, ϕ) is calibrated at e if* $\inf_{h \in \mathcal{C}_e} L_\rho^\phi(h) > \inf_{h \in \mathcal{H}_{\text{all}}} L_\rho^\phi(h)$.

The set $\mathcal{C}_e$ consists of the scores that violate e, i.e. that rank b at least as high as a. Calibration at e is therefore the analogue of classification calibration, i.e. putting the wrong ranking on e is strictly more expensive than the unconstrained optimum, so any near-optimal score must honor the pair.

Requiring this condition to hold at every pair turns out to be too strong to be necessary. Some unanimous preferences are free to satisfy. If nothing in the data pulls back against ranking a over b, a decreasing loss will simply send $h(a) - h(b)$ to $+\infty$. The comparisons that can actually obstruct Pareto optimality are those that sit on a cycle, where raising one margin forces another down. The following definition of cyclic edges makes this distinction precise.

Definition 4.2 (Support graph and cyclic edges)**.** *The support graph G has vertex set $\mathcal{X}$ and a directed edge $a \to b$ for each $(a, b) \in \text{supp}(\rho)$. A unanimous edge $e_0 = (a, b) \in E_U$ is cyclic if G contains a directed path from b back to a, equivalently if a and b lie in the same strongly connected component. Otherwise it is acyclic.*

An acyclic unanimous edge therefore has no return path. Scoring a over b contradicts no other comparison, so calibration holds automatically.

Lemma 4.1. *Suppose ϕ is nonincreasing and bounded below with $\phi(0) > \phi_\infty := \inf_{\mathbb{R}} \phi$. Then every acyclic unanimous edge $e_0 \in E_U$ satisfies* $\inf_{h \in \mathcal{C}_{e_0}} L_\rho^\phi(h) \geq \inf_{h \in \mathcal{H}_{\text{all}}} L_\rho^\phi(h) + \rho_{e_0}\big(\phi(0) - \phi_\infty\big) > \inf_{h \in \mathcal{H}_{\text{all}}} L_\rho^\phi(h)$.

In constrast, a cyclic unanimous edge has no such escape, because it competes with the comparisons on its return cycle, whose margins are forced to sum to zero. Therefore, Pareto Optimality is only feasible when (ρ, ϕ) is calibrated at every cyclic unanimous edges.

Definition 4.3. *(ρ, ϕ) is Pareto calibrated if it is calibrated at every cyclic unanimous edge $e \in E_U$.*

In the context of binary classification (Example 2.1), each x is an isolated two-label comparison, so a deterministic label $\eta(x) \in \{0, 1\}$ is an acyclic unanimous edge and PO comes for free. The interesting calibration condition is at $\eta \neq \frac{1}{2}$, which corresponds to PMC.

For the monotonicity condition that made certain losses successful, we can also show that monotonicity implies Pareto calibration.

Lemma 4.2. *If ρ is monotone at a unanimous edge $e_0 = (a, b)$ and ϕ is non-increasing, bounded below, attains its infimum on $\mathcal{C}_{e_0} = \{h \in \mathcal{H}_{\text{all}} : m_{e_0}(h) \leq 0\}$, and satisfies $\phi(\mathbb{R}_{--}) > \phi(\mathbb{R}_{++})$ with $\phi'_+(0) < 0$, then (ρ, ϕ) is Pareto calibrated at e_0.*

For strictly decreasing losses (e.g. exponential or logistic), agreeing with a unanimous edge drives margins to $+\infty$, so L_ρ^ϕ need not attain a minimizer and guarantees about minimizers are vacuous. The more useful notion is therefore asymptotic Pareto optimality, i.e. along any minimizing sequence $\{h_n\}$, every unanimous margin eventually becomes and remains strictly positive.

Definition 4.4. *A sequence $\{h_n\} \subset \mathcal{H}_{\text{all}}$ is asymptotically Pareto optimal for ρ if there exists N such that $m_e(h_n) > 0$ for every $e \in E_U$ and every $n \geq N$.*

We can then show that Pareto calibration is necessary and sufficient for this sequential guarantee.

Theorem 4.1 (Pareto calibration characterizes PO)**.** *Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is convex and nonincreasing with $\phi'_+(0) < 0$ and bounded below. Then every minimizing sequence of L_ρ^ϕ over $\mathcal{H}_{\text{all}}$ is asymptotically Pareto optimal if and only if (ρ, ϕ) is Pareto calibrated.*

If the risk also attains its infimum, the sequential guarantee applies to each minimizer as a constant minimizing sequence.

Corollary 4.1. *Under the assumptions of Theorem 4.1, if L_ρ^ϕ attains its infimum on $\mathcal{H}_{\text{all}}$ and (ρ, ϕ) is Pareto calibrated, then every minimizer of L_ρ^ϕ is Pareto optimal. Conversely, if L_ρ^ϕ attains its infimum on each $\mathcal{C}_{e_0}$ for cyclic $e_0 \in E_U$ and every minimizer of L_ρ^ϕ on $\mathcal{H}_{\text{all}}$ is Pareto optimal, then (ρ, ϕ) is Pareto calibrated.*

5 VERIFYING PARETO CALIBRATION

Pareto calibration characterizes PO, but the defining inequality compares two infima over all scores and is not easily verifiable. In this section, we give a sufficient condition for Pareto calibration that is more general than monotonicity, and is easily verifiable.

Fix a unanimous pair $e_0 = (a, b)$. The optimal-value function $\Phi_{e_0}(s) := \inf\{L_\rho^\phi(h) : h \in \mathcal{H}_{\text{all}},\ m_{e_0}(h) = s\}$ take the least risk attainable at a prescribed margin. It is finite and convex because it is obtained by minimizing a convex function subject to an affine equality. Let $d := \frac{1}{2}(\delta_a - \delta_b)$, and adding sd raises the margin on e_0 by s. If the constrained problem at margin zero has a minimizer h_0, set $u_x := h_0(x) - h_0(a) = h_0(x) - h_0(b)$ and define

$$D(e_0) \;:=\; 2\,L_\rho^{\phi\prime}(h_0; d) \;=\; 2\rho_{ab}\,\phi'(0) + \sum_{x \neq a,b} \Big[(\rho_{ax} - \rho_{bx})\,\phi'(-u_x) - (\rho_{xa} - \rho_{xb})\,\phi'(u_x)\Big]$$

This is twice the directional derivative of the risk at h_0. Convexity links the sign of the right derivative of Φ_{e_0} at zero to a calibration bound.

Lemma 5.1. *The Pareto-calibration inequality holds at e_0 if and only if $\Phi'_{e_0}(0^+) < 0$. If the constrained problem at margin zero has a minimizer h_0, then $\Phi'_{e_0}(0^+) \le \frac{1}{2}D(e_0)$.*

We next bound this directional derivative uniformly over every score with $m_{e_0}(h) = 0$.

Lemma 5.2. *For $e_0 = (a, b)$, if $\mathcal{E}(e_0) := \sum_{x \neq a,b} \big[(\rho_{bx} - \rho_{ax})_+ + (\rho_{xa} - \rho_{xb})_+\big] < 2\kappa\,\rho_{ab}$, where $G := \sup\{|g| : g \in \partial\phi(\xi),\ \xi \in \mathbb{R}\} \in (0, \infty]$, and $\kappa := |\phi'_+(0)|/G \in [0, 1]$, then the Pareto-calibration inequality holds at e_0. If this holds at every cyclic unanimous edge, then every minimizing sequence of L_ρ^ϕ is asymptotically Pareto optimal. In particular, every minimizer, when one exists, is Pareto optimal.*

Conceptually, the quantity $\mathcal{E}(e_0)$ measures how much the data around (a, b) violate monotonicity. The right-hand side $2\kappa\rho_{ab}$ is what the loss can absorb of that violation. The factor $\kappa = |\phi'_+(0)|/G$ is the slope of ϕ at the decision boundary, relative to its steepest slope anywhere. So far, our characterization optimizes over all measurable scores. However, the insights we have developed here can be carried towards cases where the score is restricted. Appendix F states and proves the corresponding characterization when the class induced by a parameterized policy.

5.1 LOSS TRUNCATION

Motivated by the goal of limiting the influence of badly misranked comparisons, we cap how fast the loss can change. In the following, we define a family of truncated losses, show that they remain classification-calibrated, and give an explicit condition under which minimizing sequences are asymptotically Pareto optimal.

Throughout this subsection, let $\phi : \mathbb{R} \to \mathbb{R}$ be C^1 and strictly decreasing, with $\phi_\infty := \inf_{\mathbb{R}} \phi > -\infty$, $\phi(\xi) \to \infty$ as $\xi \to -\infty$, and $\phi'(0) < 0$. Write $\Delta_\phi := \phi(0) - \phi_\infty > 0$. The exponential $\phi(\xi) = e^{-\xi}$ and logistic $\phi(\xi) = \log(1 + e^{-\xi})$ losses satisfy these conditions with $\Delta_\phi = 1$ and $\Delta_\phi = \log 2$, respectively.

Fix a saturator $\tau \in C^1([0, \infty), [0, 1))$ that is strictly increasing and concave, with $\tau(0) = 0$, $\tau'(t) > 0$ for every $t \ge 0$, $\tau'(0) = 1$, and $\lim_{t\to\infty} \tau(t) = 1$. Standard choices are $\tau(t) = t/(1 + t)$ and $\tau(t) = 1 - e^{-t}$. For $C > 0$, set $\phi_C(\xi) := \phi_\infty + C\,\tau\Big(\frac{\phi(\xi) - \phi_\infty}{C}\Big)$, $\Delta_C := \phi_C(0) - \phi_\infty = C\,\tau(\Delta_\phi/C)$. The cap C is the width of the open band $(\phi_\infty, \phi_\infty + C)$ into which ϕ is squeezed. A small C flattens the upper tail more strongly. A large C recovers the original loss pointwise, since $\phi_C \to \phi$ as $C \to \infty$.

Lemma 5.3. *(i) $\phi_C \in C^1$ and is strictly decreasing. (ii) $\phi_\infty < \phi_C(\xi) < \phi_\infty + C$ for every ξ, with $\phi_C(\xi) \to \phi_\infty + C$ as $\xi \to -\infty$ and $\phi_C(\xi) \to \phi_\infty$ as $\xi \to \infty$. (iii) $\phi'_C(0) = \tau'(\Delta_\phi/C)\phi'(0) < 0$.*

Although saturation bounds the range of ϕ_C, it does not in general make ϕ_C convex or make Lemma 5.2 applicable. It nevertheless preserves classification calibration.

Lemma 5.4. *For every $C > 0$, the truncated loss ϕ_C is classification-calibrated.*

Given per-edge caps $\boldsymbol{C} = (C_e)_{e\in\mathrm{supp}(\rho)}$ with $C_e > 0$, set $L_{\boldsymbol{C}}(h) := \sum_{e\in\mathrm{supp}(\rho)} \rho_e\, \phi_{C_e}(m_e(h))$ for $h \in \mathcal{H}_{\mathrm{all}}$. Write $\omega^U_{\min} := \min_{e\in E_U} \rho_e$ and $\Delta_{C_e} := \phi_{C_e}(0) - \phi_\infty$.

Truncation bounds the cost of reversing a pair, but the unanimous pairs must still be jointly realizable, so that a single score can make every unanimous margin arbitrarily large.

Lemma 5.5. *If ρ is Pareto realizable, then for every $\gamma > 0$ there exists $h^U_\gamma \in \mathcal{H}_{\mathrm{all}}$ with $m_e(h^U_\gamma) \geq \gamma$ for every $e \in E_U$.*

Thus with calibration and realizable unanimous edges, we have the following theorem.

Theorem 5.1. *Assume ρ is Pareto realizable and $E_U \neq \emptyset$. If $\sum_{e\in E_N} \rho_e C_e < \omega^U_{\min} \min_{e\in E_U} \Delta_{C_e}$ then every minimizing sequence of $L_{\boldsymbol{C}}$ over $\mathcal{H}_{\mathrm{all}}$ is asymptotically Pareto optimal. In particular, every minimizer, if one exists, is Pareto optimal.*

Truncation is not the only way to restore Lemma 5.2. Instead, one can keep ϕ unchanged and increase the unanimous comparisons until violating them is too costly. Through duality, we can show that every minimizer of the truncated risk $L_{\boldsymbol{C}}$ also minimizes a reweighted surrogate under an edgewise-optimal reweighting, and conversely. See Appendix I for further discussions.

6 Experiment

In this section, we empirically validate the effectiveness of truncating the loss and reweighting the data in preference based optimization of large language models. In this case, the score is restricted to a shared language-model policy, i.e. $\mathcal{X}$ being the space of prompt and responses (p, y), then $h_\theta(p, y) = \beta \log \frac{\pi_\theta(y|p)}{\pi_{\mathrm{ref}}(y|p)}$. We then validate whether truncation and reweighting improves the probability of satisfying Pareto optimality.

We use the human comparison split of OpenAI TL;DR (Stiennon et al., 2020). Response omparisons that share a prompt induce a finite directed graph and a single comparison measure ρ. We retain the responses for which there are at least three alternative responses under the same prompt. This leaves 442 prompts and 18,693 comparisons.

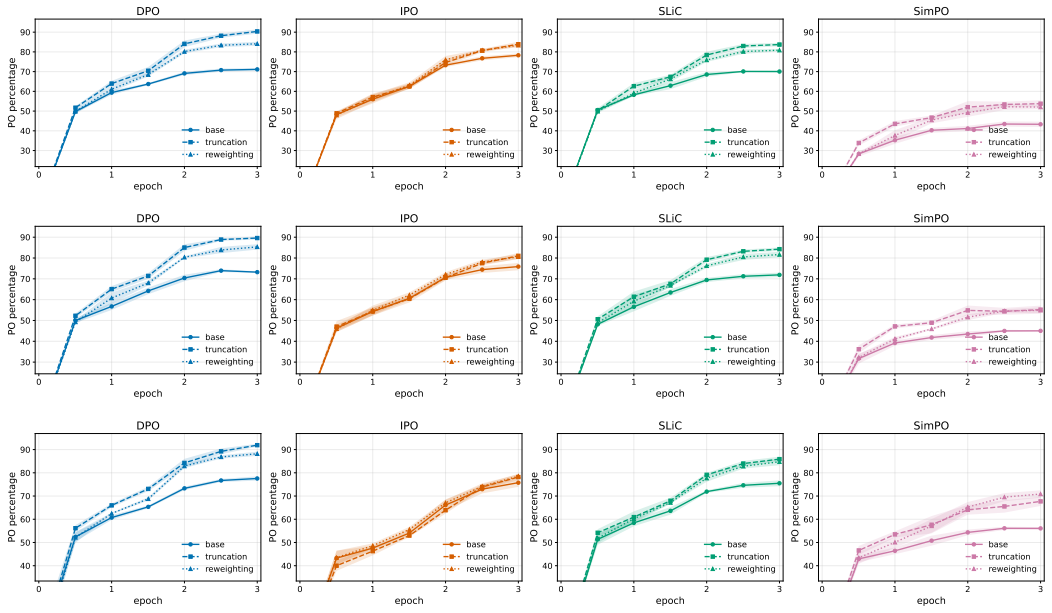

Figure 1: Percentage of satisfying Pareto Optimality (PO) on the 442 TL;DR prompts during training. Top panel with Qwen2.5-0.5B-Instruct, middle panel with Qwen2.5-3B-Instruct, and bottom panel with Zephyr SFT checkpoint of Mistral-7B.

We compare the effect of truncation and reweighting with four losses: the logistic loss of DPO (Rafailov et al., 2023), $\phi(m) = \log(1 + e^{-m})$; the squared loss of IPO (Azar et al., 2024),

$\phi(m) = (m/\beta - 1/(2\beta))^2$; the pairwise hinge of SLiC (Zhao et al., 2023), $\phi(m) = \max(0, 1-m)$; and the length-normalized logistic loss of SimPO (Meng et al., 2024). Each surrogate is run as a base objective, with the saturating truncation of Section 5 and with adaptive reweighting which is shown in Section I to be equivalent as truncation. The models are Qwen2.5-0.5B-Instruct, Qwen2.5-3B-Instruct (Qwen et al., 2025), and the Zephyr SFT checkpoint of Mistral-7B (Tunstall et al., 2023). More implementation details can be found in Appendix J. Figure 1 summarizes our empirical results, which shows truncation and adaptive reweighting improves the Pareto calibration satisfaction percentage empirically across model sizes.

7 Related Work

Classification and ranking calibration. Classification calibration studies when a tractable surrogate can replace a target loss that is hard to optimize, most classically the 0-1 loss. Early work studied binary classification (Zhang, 2004; Steinwart, 2005; Bartlett et al., 2006; Reid & Williamson, 2010). In particular, Bartlett et al. (2006) characterize calibrated surrogates and relate surrogate regret to classification regret. Tewari & Bartlett (2007) study multiclass calibration and identify inconsistent multiclass surrogates. Ramaswamy & Agarwal (2016) develop conditions for general finite loss matrices and study the prediction dimension required by convex calibrated surrogates. These results primarily concern unrestricted prediction classes; Zhang & Agarwal (2020) show that restrictions on the function class can invalidate calibration. Related work studies surrogate consistency for ranking from pairwise comparisons (Gao & Zhou, 2015; Mao et al., 2023; 2025), including bipartite ranking and related objectives (Cossock & Zhang, 2008; Ravikumar et al., 2011; Duchi et al., 2010; Calauzenes et al., 2012; Agarwal, 2014). Those settings typically assess average ranking loss, whereas we ask whether a score respects the majority and unanimous comparisons in the data.

Social choice and preference-based alignment. Social choice provides a long-standing axiomatic framework for comparing preference-aggregation rules (Arrow et al., 2010; Brandt et al., 2016). Preference-based alignment has related connections to social choice. Bradley–Terry reward fitting has been linked to Borda ranking, while methods designed for cyclic or intransitive preferences use game-theoretic solution concepts such as von Neumann winners and maximal lotteries (Christiano et al., 2017; Siththaranjan et al., 2024). In particular, Nash learning from human feedback learns a randomized policy by solving a two-player preference game rather than fitting a scalar reward whose differences explain all comparisons (Munos et al., 2024). Our work studies the latter score-based setting and we characterize when surrogate-risk minimization respects pairwise majority and unanimous comparisons.

Recent work uses social choice to analyze alignment (Conitzer et al., 2024; Dai & Fleisig, 2024; Qiu, 2024; Procaccia et al., 2025; Xiao et al., 2025; Kim et al., 2026); Gölz et al. (2026) study distortion in preference optimization. Closest to our setting, Noothigattu et al. (2020) examine maximum-likelihood estimators for random-utility models, while Ge et al. (2024) show that linear reward rules can violate PO and PMC under the Bradley–Terry model. Hollender & Kraiczy (2026) extend negative results to polynomial reward models and identify uniform directional coverage as a setting that restores PO for broad classes of loss-based rules.

8 Conclusion

We studied when minimizing surrogate margin losses on pairwise comparisons yields scores satisfying Pareto optimality (PO) and pairwise majority consistency (PMC). Under conditions like spherical symmetry, classification calibration guarantees both axioms, while it cannot ensure PO for arbitrary comparison measures. We introduced Pareto calibration, a necessary and sufficient condition on the loss and comparison measure, along with a verifiable sufficient condition for differentiable convex losses. Based on this condition, we introduce truncation and unanimous-edge reweighting, which improve empirical PO across preference objectives and model sizes.

AI USE STATEMENT

In this work, we used generative AI tools to assist in the writing and proofreading of mathematical proofs and to help implement the experimental methods. Additionally, we used generative AI tools to create and edit the experiment code and to edit the manuscript for readability. We have not used generative AI tools to generate research ideas, to discover research topics or identify gaps, or to conduct the research autonomously. We have reviewed all AI-assisted work. The authors checked the AI-assisted proofs, verified and tested the experiment code, and reviewed the polished paper. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Shivani Agarwal. Surrogate regret bounds for bipartite ranking via strongly proper losses. *The Journal of Machine Learning Research*, 15(1):1653–1674, 2014.

Kenneth J Arrow, Amartya Sen, and Kotaro Suzumura. *Handbook of social choice and welfare*, volume 2. Elsevier, 2010.

Mohammad Gheshlaghi Azar, Zhaohan Daniel Guo, Bilal Piot, Remi Munos, Mark Rowland, Michal Valko, and Daniele Calandriello. A general theoretical paradigm to understand learning from human preferences. In *International conference on artificial intelligence and statistics*, pp. 4447–4455. PMLR, 2024.

John Bartholdi III, Craig A Tovey, and Michael A Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and welfare*, 6(2):157–165, 1989.

Peter L Bartlett, Michael I Jordan, and Jon D McAuliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156, 2006.

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

Clément Calauzenes, Nicolas Usunier, and Patrick Gallinari. On the (non-) existence of convex, calibrated surrogate losses for ranking. *Advances in Neural Information Processing Systems*, 25, 2012.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.

Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H. Holliday, Bob M. Jacobs, Nathan Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey Schoelkopf, Emanuel Tewolde, and William S. Zwicker. Position: Social choice should guide AI alignment in dealing with diverse human feedback. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=w1d9DOGymR.

David Cossock and Tong Zhang. Statistical analysis of bayes optimal subset ranking. *IEEE Transactions on Information Theory*, 54(11):5140–5154, 2008.

Jessica Dai and Eve Fleisig. Mapping social choice theory to RLHF. In *ICLR 2024 Workshop on Reliable and Responsible Foundation Models*, 2024. URL https://openreview.net/forum?id=iYthn3WATc.

John C Duchi, Lester W Mackey, and Michael I Jordan. On the consistency of ranking algorithms. In *International Conference on Machine Learning*, pp. 327–334, 2010.

Wei Gao and Zhi-Hua Zhou. On the consistency of auc pairwise optimization. *International Joint Conference on Artificial Intelligence*, 2015.

Luise Ge, Daniel Halpern, Evi Micha, Ariel D Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu. Axioms for ai alignment from human feedback. *Advances in Neural Information Processing Systems*, 37:80439–80465, 2024.

Paul Gölz, Nika Haghtalab, and Kunhe Yang. Distortion of ai alignment: Does preference optimization optimize for preferences? *Advances in Neural Information Processing Systems*, 38: 25969–26007, 2026.

Alexandros Hollender and Sonja Kraiczy. Enforcing axioms for ai alignment under loss-based rules. In *The Fourteenth International Conference on Learning Representations*, 2026.

John G Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.

Kihyun Kim, Jiawei Zhang, Asuman Ozdaglar, and Pablo Parrilo. Beyond rlhf and nlhf: Population-proportional alignment under an axiomatic framework. In *International Conference on Learning Representations*, volume 2026, pp. 44621–44647, 2026.

Anqi Mao, Mehryar Mohri, and Yutao Zhong. h-consistency bounds for pairwise misranking loss surrogates. In *International conference on Machine learning*, pp. 23743–23802. PMLR, 2023.

Anqi Mao, Mehryar Mohri, and Yutao Zhong. Enhanced h-consistency bounds. In *36th International Conference on Algorithmic Learning Theory*, 2025. URL https://openreview.net/forum?id=qgnVGFJMJo.

Yu Meng, Mengzhou Xia, and Danqi Chen. Simpo: Simple preference optimization with a reference-free reward. *Advances in Neural Information Processing Systems*, 37:124198–124235, 2024.

Rémi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mésnard, Andrea Michi, et al. Nash learning from human feedback. In *International Conference on Machine Learning*. PMLR, 2024.

Ritesh Noothigattu, Dominik Peters, and Ariel D Procaccia. Axioms for learning from pairwise comparisons. *Advances in Neural Information Processing Systems*, 33:17745–17754, 2020.

Ariel D. Procaccia, Benjamin Schiffer, and Shirley Zhang. Clone-robust AI alignment. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=vbw9hSlOaG.

Tianyi Qiu. Representative social choice: From learning theory to AI alignment. In *Pluralistic Alignment Workshop at NeurIPS 2024*, 2024. URL https://openreview.net/forum?id=k106hee6I6.

Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report, 2025. URL https://arxiv.org/abs/2412.15115.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

Harish G Ramaswamy and Shivani Agarwal. Convex calibration dimension for multiclass loss matrices. *Journal of Machine Learning Research*, 17(14):1–45, 2016.

Pradeep Ravikumar, Ambuj Tewari, and Eunho Yang. On ndcg consistency of listwise ranking methods. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 618–626. JMLR Workshop and Conference Proceedings, 2011.

Mark D Reid and Robert C Williamson. Composite binary losses. *The Journal of Machine Learning Research*, 11:2387–2422, 2010.

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

Ingo Steinwart. Consistency of support vector machines and other regularized kernel classifiers. *IEEE transactions on information theory*, 51(1):128–142, 2005.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. *Advances in neural information processing systems*, 33:3008–3021, 2020.

Ambuj Tewari and Peter L. Bartlett. On the consistency of multiclass classification methods. *Journal of Machine Learning Research*, 8:1007–1025, 2007.

Lewis Tunstall, Edward Beeching, Nathan Lambert, Nazneen Rajani, Kashif Rasul, Younes Belkada, Shengyi Huang, Leandro Von Werra, Clémentine Fourrier, Nathan Habib, et al. Zephyr: Direct distillation of lm alignment. *arXiv preprint arXiv:2310.16944*, 2023.

Jiancong Xiao, Zhekun Shi, Kaizhao Liu, Qi Long, and Weijie J Su. Theoretical tensions in rlhf: Reconciling empirical success with inconsistencies in social choice theory. *arXiv preprint arXiv:2506.12350*, 2025.

Mingyuan Zhang and Shivani Agarwal. Bayes consistency vs. h-consistency: The interplay between surrogate loss functions and the scoring function class. *Advances in neural information processing systems*, 33:16927–16936, 2020.

Tong Zhang. Statistical behavior and consistency of classification methods based on convex risk minimization. *The Annals of Statistics*, 32(1):56–85, 2004.

Yao Zhao, Rishabh Joshi, Tianqi Liu, Misha Khalman, Mohammad Saleh, and Peter J Liu. Slic-hf: Sequence likelihood calibration with human feedback. *arXiv preprint arXiv:2305.10425*, 2023.

A Instances of the Framework

A.1 Multi-class Classification

Let $\mathcal{X}$ be a feature space, let $[K]$ be the label set, and take the alternative set to be $\mathcal{A}_{\mathrm{mc}} := \mathcal{X} \times [K]$. Let P be a distribution on $\mathcal{X} \times [K]$. A labeled example $(X, Y) = (x, y)$ contributes the comparisons $(x, y) \to (x, b)$ for every $b \neq y$.

Define $Q\big((x, y), A\big) := \sum_{b \neq y} \mathbf{1}_{\{((x,y),(x,b)) \in A\}}$, for every $A \subset \mathcal{A}_{\mathrm{mc}}^2$, and let $\rho_{\mathrm{mc}} := PQ$ be the induced measure on $\mathcal{A}_{\mathrm{mc}}^2$. For each fixed feature value $x \in \mathcal{X}$, among label comparisons at x one has $\rho_{\mathrm{mc},(x,a),(x,b)} = P(X = x, Y = a)$ for $a \neq b$ and $\rho_{\mathrm{mc},(x,a),(x,a)} = 0$.

Fix $x \in \mathcal{X}$ with $P(X = x) > 0$ and write $p_a(x) := P(Y = a \mid X = x)$. Then

$$\Delta_{\rho_{\mathrm{mc}}}((x, a), (x, b)) = P(X = x, Y = a) - P(X = x, Y = b) = P(X = x)\big(p_a(x) - p_b(x)\big)\,.$$

Lemma A.1. *Fix x with $P(X = x) > 0$, and let $\Gamma^\star(x) := \operatorname{argmax}_{y \in [K]} p_y(x)$ be the set of Bayes labels so that the Bayes risk is $R^\star = \mathbb{E}_X[1 - \max_y p_y(X)]$. If h satisfies pairwise majority for ρ_{mc}, then every score maximizer $\widehat{y}_h(x) \in \operatorname{argmax}_{y \in [K]} h(x, y)$ is Bayes optimal, i.e. $\widehat{y}_h(x) \in \Gamma^\star(x)$.*

Proof. Let $\widehat{y} \in \operatorname{argmax}_y h(x, y)$. Assume that $\widehat{y} \notin \Gamma^\star(x)$. Then there exists a label $a \in [K]$ such that $p_a(x) > p_{\widehat{y}}(x)$. Moreover,

$$\Delta_{\rho_{\mathrm{mc}}}((x, a), (x, \widehat{y})) = P(X = x)\big(p_a(x) - p_{\widehat{y}}(x)\big) > 0\,,$$

Pairwise majority consistency therefore gives $h(x, a) > h(x, \widehat{y})$, contradicting the choice of $\widehat{y}$ as a score maximizer. Thus every score maximizer belongs to $\Gamma^\star(x)$.

Conversely, if a is the unique maximizer of $p_y(x)$, then $p_a(x) > p_b(x)$ for every $b \neq a$, so pairwise majority gives $h(x, a) > h(x, b)$ for every $b \neq a$. Hence a is the unique score maximizer. □

A.2 Pairwise Ranking

Let X, X' be two items and let $Y = +1$ mean that X' should be ranked above X. Conditional on the pair (x, x'), write $\eta(x, x') := \mathbb{P}(Y = +1 \mid X = x, X' = x')$. The aggregate comparison measure ρ_{rank} on $\mathcal{X}^2$ is

$$\rho_{\mathrm{rank},x',x} = \pi(x, x')\,\eta(x, x'), \qquad \rho_{\mathrm{rank},x,x'} = \pi(x, x')\big(1 - \eta(x, x')\big),$$

where π is the pair-sampling distribution. Then $\Delta_{\rho_{\mathrm{rank}}}(x', x) = \pi(x, x')\big(2\eta(x, x') - 1\big)$.

A.3 Linear Social Choice

In linear social choice Ge et al. (2024); Hollender & Kraiczy (2026) the alternatives are points in a Euclidean embedding space and both the voters and the score are linear. Here the alternative set is a finite candidate set $\mathcal{X} \subseteq \mathbb{R}^d$, as in Section 2. The spherical continuum model in Section 3 instead integrates against a comparison-direction measure. There are n voters, where voter i carries a direction $v_i \in \mathbb{R}^d \setminus \{0\}$ and a weight $w_i > 0$, and ranks a above b when $\langle v_i, a - b\rangle > 0$. Voter i decides the oriented pairs in the open halfspace it induces,

$$T(v_i) := \big\{(a, b) \in \mathcal{X}^2 : \langle v_i, a - b\rangle > 0\big\},$$

and the induced comparison measure sums each voter's decisive pairs with its weight,

$$\rho_{\mathrm{lin}} := \sum_{i=1}^{n} w_i \sum_{(a,b) \in T(v_i)} \delta_{(a,b)}, \qquad \rho_{\mathrm{lin},ab} = \sum_{i:\,\langle v_i, a-b\rangle > 0} w_i.$$

The linear social choice restricts h to the linear class $\mathcal{H}_{\mathrm{lin}} := \{x \mapsto \langle \theta, x\rangle : \theta \in \mathbb{R}^d\} \subseteq \mathcal{H}_{\mathrm{all}}$. For $h = \langle \theta, \cdot\rangle$ the margin is $m_{ab}(h) = \langle \theta, a - b\rangle$, so the risk of Equation 1 becomes the weighted total loss of the reward θ against the voter comparisons,

$$L^{\phi}_{\rho_{\mathrm{lin}}}\big(\langle \theta, \cdot\rangle\big) = \sum_{i=1}^{n} w_i \sum_{(a,b) \in T(v_i)} \phi\big(\langle \theta, a - b\rangle\big).$$

As $\Delta_{\rho_{\mathrm{lin}}}(a,b) = \sum_{i:\,\langle v_i, a-b\rangle>0} w_i - \sum_{i:\,\langle v_i, b-a\rangle>0} w_i$, (a,b) is unanimous, exactly when every voter prefers a to b, that is $\langle v_i, a-b\rangle > 0$ for all i.

B Direct Preference Optimization

Direct preference optimization (DPO) keeps the same alternative set $\mathcal{A}_{\mathrm{BT}}$, comparison measure ρ_{BT}, and logistic loss ϕ as Example 2.2, but replaces the free reward by an implicit reward tied to a policy. Fixing a reference policy π_{ref} and inverse temperature $\beta > 0$, the score is restricted to the class $h_\theta(x,y) := \beta \log \frac{\pi_\theta(y|x)}{\pi_{\mathrm{ref}}(y|x)}$. Because the margin depends only on score differences, the prompt-only normalizer $\beta \log Z(x)$ of the DPO reparameterization cancels, and $L^{\phi}_{\rho_{\mathrm{BT}}}(h_\theta)$ is exactly the DPO objective

$$\mathbb{E}_{(X,Y_w,Y_l)\sim P}\Big[\log\Big(1+\exp\big(-\beta\log\tfrac{\pi_\theta(Y_w|X)}{\pi_{\mathrm{ref}}(Y_w|X)} + \beta\log\tfrac{\pi_\theta(Y_l|X)}{\pi_{\mathrm{ref}}(Y_l|X)}\big)\Big)\Big]$$

C Proofs for the 0-1 Loss and Kemeny Rule

Proposition 2.1 (0-1 risk equals generalized Kemeny)**.** *For every ρ and every score h, and every strict total order σ that linearly extends the strict relation induced by h, $K_\rho(\sigma) \le L^{0\text{-}1}_\rho(h)$ with equality when h is injective. Consequently, $\inf_{h:\mathcal{X}\to\mathbb{R}} L^{0\text{-}1}_\rho(h) = \min_\sigma K_\rho(\sigma)$ where the minimum on the right ranges over strict total orders on $\mathcal{X}$.*

Proof. If h is injective, let σ_h be the strict total order with $a \succ_{\sigma_h} b$ iff $h(a) > h(b)$. Fix $a \neq b$. If $h(a) > h(b)$, then $a \succ_{\sigma_h} b$, so $b \not\succ_{\sigma_h} a$ and $\mathbf{1}_{h(a)\le h(b)} = 0 = \mathbf{1}_{b\succ_{\sigma_h} a}$. If $h(a) < h(b)$, then $b \succ_{\sigma_h} a$ and $\mathbf{1}_{h(a)\le h(b)} = 1 = \mathbf{1}_{b\succ_{\sigma_h} a}$. Thus $\mathbf{1}_{h(a)\le h(b)} = \mathbf{1}_{b\succ_{\sigma_h} a}$ for every $a \neq b$. Summing over ordered pairs and using $\rho_{aa} = 0$ gives $L^{0\text{-}1}_\rho(h) = K_\rho(\sigma_h)$.

Fix any score h and a strict total order σ that linearly extends its strict score relation, i.e. $h(a) > h(b)$ implies $a \succ_\sigma b$. For each $a \neq b$, if $h(a) > h(b)$, then $a \succ_\sigma b$, so $\mathbf{1}_{b\succ_\sigma a} = 0 = \mathbf{1}_{h(a)\le h(b)}$. If $h(a) \le h(b)$, then $\mathbf{1}_{h(a)\le h(b)} = 1 \ge \mathbf{1}_{b\succ_\sigma a}$. Hence $\mathbf{1}_{b\succ_\sigma a} \le \mathbf{1}_{h(a)\le h(b)}$ for every $a \neq b$.

Multiplying by $\rho_{ab} \ge 0$ and summing yields $K_\rho(\sigma) \le \sum_{a\neq b} \rho_{ab}\,\mathbf{1}_{h(a)\le h(b)} = L^{0\text{-}1}_\rho(h)$, which is $K_\rho(\sigma) \le L^{0\text{-}1}_\rho(h)$. When h is injective, taking $\sigma = \sigma_h$ turns this into an equality.

Conversely, given any strict total order σ, enumerate $\mathcal{X}$ as $x_1 \succ_\sigma \cdots \succ_\sigma x_m$ and define $h_\sigma(x_k) := m - k$. Then h_σ is injective and $h_\sigma(x_i) > h_\sigma(x_j)$ iff $i < j$ iff $x_i \succ_\sigma x_j$, so $\sigma_{h_\sigma} = \sigma$. Injectivity therefore gives $L^{0\text{-}1}_\rho(h_\sigma) = K_\rho(\sigma)$, and hence $\inf_h L^{0\text{-}1}_\rho(h) \le \min_\sigma K_\rho(\sigma)$.

Every score h admits a strict extension σ obtained by breaking ties arbitrarily, and then we have $K_\rho(\sigma) \le L^{0\text{-}1}_\rho(h)$. Taking the minimum over σ on the left yields $\min_\sigma K_\rho(\sigma) \le L^{0\text{-}1}_\rho(h)$ for every h, so $\min_\sigma K_\rho(\sigma) \le \inf_h L^{0\text{-}1}_\rho(h)$. Combining the two inequalities proves $\inf_h L^{0\text{-}1}_\rho(h) = \min_\sigma K_\rho(\sigma)$. □

Theorem 2.1. *Fix a pair (a,b) such that $\Delta_\rho(a,b) > 0$. If ρ is monotone at (a,b), then every Kemeny minimizer $\sigma^\star$ ranks $a \succ_{\sigma^\star} b$, and every minimizer $h^\star$ of $L^{0\text{-}1}_\rho$ over all scores satisfies $h^\star(a) > h^\star(b)$. Consequently, 1) if ρ is monotone at every strict majority pair, then every Kemeny minimizer and every 0-1 minimizer satisfies pairwise majority consistency. 2) if ρ is monotone at every unanimous pair (every (a,b) with $\rho_{ba} = 0 < \rho_{ab}$), then every such minimizer is Pareto optimal.*

Proof. Fix a pair (a,b) such that $\Delta_\rho(a,b) > 0$ and assume there exists a Kemeny minimizer σ with $b \succ_\sigma a$. Let

$$C := \{c \in \mathcal{X} : b \succ_\sigma c \succ_\sigma a\}$$

be the set of alternatives strictly between b and a in σ. Let σ' be obtained from σ by interchanging the positions of a and b and leaving every other alternative in place, i.e. a moves to the position formerly held by b and b to the slot formerly held by a, while each $c \in C$ keeps its position. The only pairs whose orientation changes are $\{a,b\}$ and, for each $c \in C$, the pairs $\{a,c\}$ and $\{b,c\}$.

The change in discordance from σ to σ' for $\{a, b\}$ is $\rho_{ba} - \rho_{ab} = -\Delta_\rho(a, b)$. For each $c \in C$ we had $b \succ_\sigma c \succ_\sigma a$ and now have $a \succ_{\sigma'} c \succ_{\sigma'} b$, so the two affected pairs contribute

$$(\rho_{ca} - \rho_{ac}) + (\rho_{bc} - \rho_{cb}) = \big(\rho_{ca} - \rho_{cb}\big) + \big(\rho_{bc} - \rho_{ac}\big).$$

Taking the summation gives

$$K_\rho(\sigma') - K_\rho(\sigma) = -\Delta_\rho(a, b) - \sum_{c \in C} \Big[\big(\rho_{cb} - \rho_{ca}\big) + \big(\rho_{ac} - \rho_{bc}\big)\Big] \leq -\Delta_\rho(a, b) < 0\,,$$

where the inequality uses monotonicity at (a, b). Specifically, monotonicity gives $\rho_{cb} \geq \rho_{ca}$ and $\rho_{ac} \geq \rho_{bc}$ for every $c \in C$.

This contradicts the optimality of σ. Hence $a \succ_\sigma b$ in every Kemeny minimizer. By Proposition 2.1, we have every minimizer $h^\star$ of $L_\rho^{0\text{-}1}$ satisfies $h^\star(a) > h^\star(b)$.

The remaining two claims follow by applying the pairwise statement to each majority pair, and to each unanimous pair. □

Theorem 2.2. *Fix a pair (a, b) with $\Delta_\rho(a, b) > 0$ and suppose ρ is monotone at (a, b). Let $\phi : \mathbb{R} \to \mathbb{R}$ be nonincreasing with $\phi(\mathbb{R}_{--}) > \phi(\mathbb{R}_{++})$, and suppose L_ρ^ϕ attains its minimum over all scores. Then every minimizer $h^\star$ satisfies $h^\star(a) \geq h^\star(b)$. If, in addition, ϕ is differentiable on $\mathbb{R}$ with $\phi'(0) < 0$ then $h^\star(a) > h^\star(b)$.*

Proof. Let h be a minimizer of L_ρ^ϕ. First suppose $h(a) < h(b)$, and let $\bar{h}$ be obtained from h by swapping only the two scores $h(a)$ and $h(b)$. Let $r = h(b) - h(a) > 0$. The contribution of the pair $\{a, b\}$ changes by

$$\rho_{ab}\big[\phi(r) - \phi(-r)\big] + \rho_{ba}\big[\phi(-r) - \phi(r)\big] = \Delta_\rho(a, b)\big[\phi(r) - \phi(-r)\big].$$

Since $-r \in \mathbb{R}_{--}$ and $r \in \mathbb{R}_{++}$, the assumption $\phi(\mathbb{R}_{--}) > \phi(\mathbb{R}_{++})$ gives $\phi(-r) > \phi(r)$, so this change is strictly negative.

For any $c \in \mathcal{X} \setminus \{a, b\}$, the terms involving c change by

$$(\rho_{ca} - \rho_{cb})\Big[\phi\big(h(c) - h(b)\big) - \phi\big(h(c) - h(a)\big)\Big] + (\rho_{ac} - \rho_{bc})\Big[\phi\big(h(b) - h(c)\big) - \phi\big(h(a) - h(c)\big)\Big].$$

Because $h(a) < h(b)$ and ϕ is nonincreasing, the first bracket is nonnegative and the second bracket is nonpositive. Monotonicity at (a, b) gives $\rho_{cb} \geq \rho_{ca}$ and $\rho_{ac} \geq \rho_{bc}$, so both terms are nonpositive. Hence $L_\rho^\phi(\bar{h}) < L_\rho^\phi(h)$, contradicting the optimality of h. Therefore no minimizer can satisfy $h(a) < h(b)$, and every minimizer satisfies $h(a) \geq h(b)$.

Now we rule out ties: suppose $h(a) = h(b) = s$. For $t > 0$, define h_t by $h_t(a) = s{+}t$, $h_t(b) = s{-}t$, $h_t(c) = h(c)$ for $c \notin \{a, b\}$. If ϕ is differentiable, and $\phi'(0) < 0$,

$$\frac{d}{dt} L_\rho^\phi(h_t)\bigg|_{t=0+} = 2\Delta_\rho(a, b)\phi'(0) + \sum_{c \neq a,b} (\rho_{cb} - \rho_{ca})\phi'\big(h(c) - s\big) + \sum_{c \neq a,b} (\rho_{ac} - \rho_{bc})\phi'\big(s - h(c)\big).$$

The first term is strictly negative because $\Delta_\rho(a, b) > 0$ and $\phi'(0) < 0$. The two sums are nonpositive because monotonicity gives $\rho_{cb} \geq \rho_{ca}$ and $\rho_{ac} \geq \rho_{bc}$, while nonincreasing differentiable losses satisfy $\phi' \leq 0$. Thus the one-sided derivative is strictly negative, so $L_\rho^\phi(h_t) < L_\rho^\phi(h)$ for all sufficiently small $t > 0$. This contradicts optimality. Hence every surrogate minimizer satisfies $h(a) > h(b)$. □

D Proof of the Negative Result

Theorem 3.1. *Let ϕ be continuous, convex, and coercive with a unique minimizer. Then there exists a PO-realizable measure ρ such that every minimizer of L_ρ^ϕ over $\mathcal{H}_{\text{all}}$ violates Pareto optimality.*

Proof. Fix ϕ continuous, convex, and coercive with unique minimizer $m^\star$.

If $m^\star \le 0$, take $\mathcal{X} = \{a, b\}$ and $\rho_{ab} = 1$, with zero mass elsewhere. The unanimous graph consists of the single edge $a \to b$ and is acyclic. Minimizing the risk is equivalent to minimizing $\phi(m)$ over $m = h(a) - h(b)$, whose unique solution is $m = m^\star \le 0$. Thus every minimizer violates Pareto optimality.

Suppose now that $m^\star > 0$. Fix $\varepsilon > 0$ and choose

$$K > \varepsilon \, \frac{\phi(-2m^\star) - \phi(m^\star)}{\phi(0) - \phi(m^\star)}.$$

Both numerator and denominator are positive because $m^\star$ is the unique minimizer and $-2m^\star, 0 \neq m^\star$. On three alternatives, first define

$$F_0(x, y) := 2\varepsilon\phi(x) + K\phi(y) + K\phi(-(x + y)),$$

where $x = h(a) - h(b)$ and $y = h(b) - h(c)$. Convexity gives

$$\phi(y) + \phi(-(x + y)) \ge 2\phi(-x/2).$$

Consequently, whenever $x \ge 0$,

$$F_0(x, y) \ge 2\varepsilon\phi(m^\star) + 2K\phi(0).$$

At $z_0 := (-2m^\star, m^\star)$,

$$F_0(z_0) = 2\varepsilon\phi(-2m^\star) + 2K\phi(m^\star).$$

The choice of K makes the gap

$$g := 2\varepsilon\phi(m^\star) + 2K\phi(0) - F_0(z_0)$$

strictly positive.

Now add reverse mass $\delta > 0$ to the two return edges and set

$$\rho_{ab} = 2\varepsilon, \qquad \rho_{bc} = \rho_{ca} = K, \qquad \rho_{cb} = \rho_{ac} = \delta,$$

with zero mass elsewhere. The only unanimous edge is $a \to b$, so ρ is Pareto realizable. Its risk is

$$F_\delta(x, y) = F_0(x, y) + \delta\big[\phi(-y) + \phi(x + y)\big].$$

For every (x, y), the bracketed term is at least $2\phi(m^\star)$, whereas at z_0 it equals $2\phi(-m^\star)$. Choose

$$0 < \delta < \frac{g}{2\big(\phi(-m^\star) - \phi(m^\star)\big)}.$$

Then every (x, y) with $x \ge 0$ satisfies

$$F_\delta(x, y) \ge F_0(z_0) + g + 2\delta\phi(m^\star) > F_0(z_0) + 2\delta\phi(-m^\star) = F_\delta(z_0).$$

Finally, F_δ is continuous and coercive, so it attains its minimum. The strict comparison above shows that every minimizer has $x < 0$ and therefore violates Pareto optimality on the sole unanimous edge $a \to b$. □

E Linear Social Choice and Symmetric Data

Lemma 3.2. *For $x \in \mathbb{S}^{d-1}$, let $n_+(x) := |\{i : \langle v_i, x\rangle > 0\}|$ and $\eta(x) := n_+(x)/n$. Then for every $\theta \in \mathbb{R}^d$, $L^\phi_{\mathrm{unif}}(\theta) = \frac{n}{2} \int_{\mathbb{S}^{d-1}} C^\phi_{\eta(x)}\big(\langle\theta, x\rangle\big)\, d\sigma(x)$.*

Proof. The set $N := \{x \in \mathbb{S}^{d-1} : \langle v_i, x\rangle = 0 \text{ for some } i\}$ is a finite union of great subspheres and is σ-null. Exchanging sum and integral in Equation **??**,

$$L^\phi_{\mathrm{unif}}(\theta) = \int_{\mathbb{S}^{d-1}} n_+(x)\, \phi(\langle\theta, x\rangle)\, d\sigma(x)$$

up to the null set N. Since σ is invariant under $x \mapsto -x$,

$$\int_{\mathbb{S}^{d-1}} n_+(x)\, \phi(\langle\theta, x\rangle)\, d\sigma(x) = \int_{\mathbb{S}^{d-1}} n_+(-x)\, \phi(-\langle\theta, x\rangle)\, d\sigma(x).$$

Averaging the two expressions and using $n_+(x) + n_+(-x) = n$ off N,

$$L^\phi_{\mathrm{unif}}(\theta) = \frac{1}{2} \int_{\mathbb{S}^{d-1}} \Big[n_+(x)\, \phi(\langle\theta, x\rangle) + \big(n - n_+(x)\big)\, \phi(-\langle\theta, x\rangle)\Big] d\sigma(x) = \frac{n}{2} \int_{\mathbb{S}^{d-1}} C^\phi_{\eta(x)}(\langle\theta, x\rangle)\, d\sigma(x).$$

□

Lemma 3.3. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be classification-calibrated, bounded below, and lower semicontinuous, and assume L^ϕ_ρ attains its minimum on $\mathcal{H}_{\mathrm{all}} := \{h : \mathbb{S}^{d-1} \to \mathbb{R}\}$. Then every minimizer $h^\star$ satisfies pairwise majority consistency for σ-almost every $x \in \mathbb{S}^{d-1}$.*

Proof. For $h \in \mathcal{H}_{\mathrm{all}}$, write the pairwise margin $m_h(x) := h(x) - h(-x)$, so $m_h(-x) = -m_h(x)$. Since σ is uniform, $n_+(-x) = n - n_+(x)$ off N, and $m_h(-x) = -m_h(x)$, reflecting $x \mapsto -x$ in the integral above and averaging yields

$$L^\phi_\rho(h) = \frac{n}{2} \int_{\mathbb{S}}^{d-1} C^\phi_{\eta(x)}\big(m_h(x)\big)\, d\sigma(x)\,, \tag{2}$$

where $C^\phi_\eta(\alpha) = \eta\,\phi(\alpha) + (1-\eta)\,\phi(-\alpha)$ is the binary conditional risk from before Definition 3.1.

Suppose $h^\star$ minimizes L^ϕ_ρ but violates PMC on a set $S \subseteq \mathbb{S}^{d-1}$ of positive σ-measure, where each $x \in S$ has $\eta(x) \neq \frac{1}{2}$. Then for $x \in S$ the margin $m_{h^\star}(x)$ has the wrong sign relative to $2\eta(x) - 1$, so it cannot minimize $\alpha \mapsto C^\phi_{\eta(x)}(\alpha)$ once ϕ is classification-calibrated; in fact $C^\phi_{\eta(x)}(m_{h^\star}(x)) \geq H^{\phi,-}(\eta(x)) > H^\phi(\eta(x))$. Define an odd function m by choosing a minimizer of $C^\phi_{\eta(y)}$ at each y and setting $m(-y) := -m(y)$. This is consistent because $C^\phi_{1-\eta}(-\alpha) = C^\phi_\eta(\alpha)$. Pick $\widetilde{h} \in \mathcal{H}_{\mathrm{all}}$ with $m_{\widetilde{h}} = m$. The integrand in Equation 2 is strictly smaller on S and no larger elsewhere, so $L^\phi_\rho(\widetilde{h}) < L^\phi_\rho(h^\star)$, a contradiction. Hence every minimizer is pointwise optimal for σ-a.e. x, i.e. $m_{h^\star}(x) \in \operatorname{argmin} C^\phi_{\eta(x)}$ for a.e. x with $\eta(x) \neq \frac{1}{2}$. This thus forces $\mathbf{sign}(m_{h^\star}(x)) = \mathbf{sign}(2\eta(x) - 1)$ for a.e. such x, which is PMC. □

E.1 Spherically Symmetric Case

We now extends the uniform-data results for linear social choice in two directions. We now allow (i) the voters to be drawn from an arbitrary distribution ν on the unit sphere, and (ii) the comparison directions to follow any spherically symmetric law μ on $\mathbb{R}^d$ (e.g. uniform on the sphere or a standard Gaussian), which sharpens the special case of Hollender & Kraiczy (2026).

Concretely, draw a comparison direction $\mathbf{X} \sim \mu$ and a voter $V \sim \nu$; the voter is decisive for the orientation

$$Y = \mathbf{sign}(\langle \mathbf{X}, V\rangle), \qquad V \sim \nu, \tag{3}$$

preferring $\mathbf{X}$ over its antipode when $Y = +1$. The surrogate risk of a linear reward $\boldsymbol{\theta}$ is then

$$\min_{\boldsymbol{\theta}}\ \mathbf{E}[\phi(Y\langle \mathbf{X}, \boldsymbol{\theta}\rangle)] =: \mathbf{E}[R(V, \boldsymbol{\theta})], \qquad R(\boldsymbol{v}, \boldsymbol{\theta}) := \mathbf{E}[\phi(\mathbf{sign}(\langle \mathbf{X}, \boldsymbol{v}\rangle)\langle \mathbf{X}, \boldsymbol{\theta}\rangle)], \tag{4}$$

the continuous counterpart of the linear social choice risk L^ϕ_ρ. Write $K := \mathrm{cl\,cone}(\operatorname{supp}\nu)$ for the closed convex cone generated by the voter directions. Membership of the optimal score direction in K is then Pareto optimality, i.e. the score is a nonnegative combination of voter directions, hence agrees with every direction on which the voters unanimously agree.

Theorem E.1 (Spherically symmetric case)**.** *Suppose $d \geq 2$, the distribution μ of comparison directions is spherically symmetric and not concentrated at the origin, and $Y = \mathbf{sign}(\langle \mathbf{X}, V\rangle)$ is the orientation preferred by a voter $V \sim \nu$ supported on the unit sphere and not a single point. Let ϕ be convex and nonincreasing with $c := \inf \phi > -\infty$ and right derivative $\phi'_+(0) < 0$. Then the risk $\boldsymbol{\theta} \mapsto \mathbf{E}[\phi(Y\langle \mathbf{X}, \boldsymbol{\theta}\rangle)]$ is coercive and attains its minimum, and every minimizer $\boldsymbol{\theta}^\star$ is nonzero with direction $\mathbf{u}^\star := \boldsymbol{\theta}^\star / \|\boldsymbol{\theta}^\star\|_2 \in K$.*

Proof. Separate the length $r := \|\boldsymbol{\theta}\|_2$ and direction $\mathbf{u} := \boldsymbol{\theta}/r$, and set $\phi_r := \phi(r\,\cdot)$. By spherical symmetry of $\mathbf{X}$,

$$R(\boldsymbol{v}, \mathbf{u}) = \mathbf{E}[\phi(r\,\mathbf{sign}(\langle \mathbf{X}, \boldsymbol{v}\rangle)\langle \mathbf{X}, \mathbf{u}\rangle)] = \mathbf{E}[\phi_r(\mathbf{sign}(\langle \mathbf{X}, \boldsymbol{v}\rangle)\langle \mathbf{X}, \mathbf{u}\rangle)] = h_r(\langle \boldsymbol{v}, \mathbf{u}\rangle), \tag{5}$$

i.e. the expectation depends on $\boldsymbol{v}, \mathbf{u}$ only through $\rho := \langle \boldsymbol{v}, \mathbf{u}\rangle$.

Writing $A := \langle \mathbf{X}, \boldsymbol{v} \rangle$ and $B := \langle \mathbf{X}, \boldsymbol{v}^\perp \rangle$ with $\boldsymbol{v}^\perp := \frac{1}{\sqrt{1-\rho^2}}(\mathbf{u} - \rho \boldsymbol{v})$,

$$h_r(\rho) = \mathbf{E}[\phi_r(\rho|A| + \sqrt{1-\rho^2}\ \mathbf{sign}(A) \cdot B)]. \tag{6}$$

Conditioned on $|A| = a \geq 0$, the variables B and $-B$ have the same conditional law, so it suffices to study, for $a, b \geq 0$,

$$g(\rho, s) := \phi_r(\rho a + sb) + \phi_r(\rho a - sb), \qquad s := \sqrt{1-\rho^2}. \tag{7}$$

Since ϕ_r is decreasing and convex and $a \geq 0$, g is nonincreasing in ρ; it is also nondecreasing in $s \geq 0$, because for $s_1 \geq s_2 \geq 0$ convexity of ϕ_r gives

$$\phi_r(\rho a + s_1 b) + \phi_r(\rho a - s_1 b) \geq \phi_r(\rho a + s_2 b) + \phi_r(\rho a - s_2 b). \tag{8}$$

As ρ increases, ρa increases while $s = \sqrt{1-\rho^2}$ decreases, and both effects lower g; averaging over A, B shows h_r is nonincreasing in ρ.

The objective equation 4 becomes

$$\min_{\boldsymbol{\theta}}\ \mathbf{E}[\phi(Y\langle \mathbf{X}, \boldsymbol{\theta} \rangle)] = \min_{r \geq 0} \min_{\|\mathbf{u}\|_2 = 1}\ \mathbf{E}[h_r(\langle V, \mathbf{u} \rangle)], \qquad V \sim \nu. \tag{9}$$

Fix $r > 0$ and suppose the inner minimizer satisfies $\mathbf{u} \notin K$. Consider $\tilde{\mathbf{u}} := \mathrm{Proj}_K(\mathbf{u}) + \eta \boldsymbol{v}^*$ for some $\boldsymbol{v}^* \in \mathbf{int}(K^*)$,[1] where K^* is the dual cone. For any $\eta > 0$ and any $V \in \operatorname{supp} \nu \subseteq K$,

$$\langle V, \tilde{\mathbf{u}} \rangle = \langle V, \mathrm{Proj}_K(\mathbf{u}) \rangle + \eta \langle V, \boldsymbol{v}^* \rangle > \langle V, \mathrm{Proj}_K(\mathbf{u}) \rangle \geq \langle V, \mathbf{u} \rangle, \tag{10}$$

$$\|\tilde{\mathbf{u}}\|_2^2 = \underbrace{\|\mathrm{Proj}_K(\mathbf{u})\|_2^2}_{< \|\mathbf{u}\|_2^2} + 2\eta \underbrace{\langle \mathrm{Proj}_K(\mathbf{u}), \boldsymbol{v}^* \rangle}_{> 0} + \eta^2 \|\boldsymbol{v}^*\|_2^2. \tag{11}$$

Hence we may choose $\eta > 0$ to restore the norm, $\|\tilde{\mathbf{u}}\|_2 = \|\mathbf{u}\|_2 = 1$, while keeping $\langle V, \tilde{\mathbf{u}} \rangle > \langle V, \mathbf{u} \rangle$ for every $V \in \operatorname{supp} \nu$. Since h_r is nonincreasing, $\mathbf{E}[h_r(\langle V, \tilde{\mathbf{u}} \rangle)] \leq \mathbf{E}[h_r(\langle V, \mathbf{u} \rangle)]$, so a minimizer with $\mathbf{u} \in K$ exists.

Under $d \geq 2$ with $\mathbf{X}$ not concentrated at the origin and $\phi'_+(0) < 0$, the monotonicity of h_r is strict: fixing a shell of $\|\mathbf{X}\|_2$ of positive mass and taking $A = \langle \mathbf{X}, \boldsymbol{v} \rangle$ near the origin (with $B = 0$) realizes the strictly decreasing part of ϕ around 0. Then $\mathbf{E}[h_r(\langle V, \tilde{\mathbf{u}} \rangle)] < \mathbf{E}[h_r(\langle V, \mathbf{u} \rangle)]$ whenever $\mathbf{u} \notin K$, so every minimizer direction lies in K.

Let $m(\mathbf{u}) := \mathbf{E}[(-Y\langle \mathbf{X}, \mathbf{u} \rangle)_+]$, which is lower semicontinuous in $\mathbf{u}$, and set $m := \min_{\|\mathbf{u}\|_2 = 1} m(\mathbf{u}) \geq 0$. If $m = 0$, then $Y\langle \mathbf{X}, \mathbf{u} \rangle \geq 0$ almost surely for some $\mathbf{u}$, i.e. $Y = \mathbf{sign}(\langle \mathbf{X}, \mathbf{u} \rangle)$ a.s., which is impossible unless $\mathrm{supp}(\nu)$ is a single point on the sphere. Hence $m > 0$. Fix a subgradient $g \in \partial\phi(0)$; since $\phi'_+(0) < 0$ we have $g < 0$, and convexity gives $\phi(t) \geq \phi(0) + gt$ for all t. Writing $\boldsymbol{\theta}_n = r_n \mathbf{u}_n$ with $r_n = \|\boldsymbol{\theta}_n\|_2$ and splitting on the sign of $Y\langle \mathbf{X}, \mathbf{u}_n \rangle$,

$$\mathbf{E}[\phi(Y\langle \mathbf{X}, r_n \mathbf{u}_n \rangle)] \geq \mathbf{E}\big[(\phi(0) + g\, Y\langle \mathbf{X}, r_n \mathbf{u}_n \rangle) \mathbf{1}_{Y\langle \mathbf{X}, \mathbf{u}_n \rangle \leq 0}\big] + c\ \Pr(Y\langle \mathbf{X}, \mathbf{u}_n \rangle > 0) \tag{12}$$

$$\geq \min(\phi(0), c) + (-g)\, r_n\, \mathbf{E}[(-Y\langle \mathbf{X}, \mathbf{u}_n \rangle)_+] \tag{13}$$

$$\geq \min(\phi(0), c) + (-g)\, r_n\, m, \tag{14}$$

using $c = \inf \phi$ finite and $m(\mathbf{u}_n) \geq m$. Since $m > 0$ and $g < 0$, this lower bound tends to ∞ as $r_n \to \infty$, so the risk is coercive; being lower semicontinuous, it attains its minimum. Finally $\boldsymbol{\theta}^\star \neq 0$: because $g < 0$, moving from the origin along any descent direction strictly decreases the risk, so no minimizer is the zero vector. □

Remark E.1. *This improves the uniform-data guarantee of Hollender & Kraiczy (2026) in three ways:*

- *The distribution μ of comparison directions need only be spherically symmetric rather than uniform on the sphere, achieved by removing every explicit computation that depends on μ.*
- *The voters may be drawn from any nontrivial distribution ν on the sphere rather than a fixed finite set. Through the hemispherical transform, the induced preferred orientation can be any odd function of the direction $\mathbf{u}$.*
- *We drop differentiability of ϕ, requiring only $\phi'_+(0) < 0$; in particular the hinge loss is covered.*

[1] The interior $\mathbf{int}(K^*)$ is taken in the span of K, which is nonempty iff K is pointed, *i.e.* $K \cap (-K) = \{0\}$.

F Calibration under a Restricted Score Class

The preceding characterization optimizes over all measurable scores. In applications, however, the score belongs to a restricted class $\mathcal{H} \subseteq \mathcal{H}_{\mathrm{all}}$. For example, DPO uses the policy-induced class

$$\mathcal{H}_\Pi := \left\{ (x, y) \mapsto \beta \log \frac{\pi_\theta(y \mid x)}{\pi_{\mathrm{ref}}(y \mid x)} : \theta \in \Theta \right\}.$$

Bayes calibration over $\mathcal{H}_{\mathrm{all}}$ need not imply consistency over $\mathcal{H}$. Coupling among the margins realizable by $\mathcal{H}$ may make a unanimous comparison incompatible with low surrogate risk Zhang & Agarwal (2020). This issue can arise even for an acyclic unanimous edge because the cut-shift argument of Lemma 4.1 need not remain inside $\mathcal{H}$.

For a nonempty class $\mathcal{H} \subseteq \mathcal{H}_{\mathrm{all}}$, define its calibration gap at a unanimous edge e by $\delta_{\mathcal{H}}(e) := \inf_{\substack{h \in \mathcal{H} \\ m_e(h) \le 0}} L_\rho^\phi(h) - \inf_{h \in \mathcal{H}} L_\rho^\phi(h)$, where the infimum of the empty set is $+\infty$. We call $(\rho, \phi, \mathcal{H})$ *$\mathcal{H}$-Pareto calibrated* if $\delta_{\mathcal{H}}(e) > 0$ for every $e \in E_U$. In contrast to Pareto calibration over $\mathcal{H}_{\mathrm{all}}$, the condition ranges over all unanimous edges rather than only cyclic ones.

Theorem F.1 ($\mathcal{H}$-Pareto calibration characterizes restricted-class PO)**.** *Assume $\inf_{h \in \mathcal{H}} L_\rho^\phi(h)$ is finite. Every sequence $\{h_n\} \subseteq \mathcal{H}$ satisfying $L_\rho^\phi(h_n) \longrightarrow \inf_{h \in \mathcal{H}} L_\rho^\phi(h)$ is asymptotically Pareto optimal if and only if $(\rho, \phi, \mathcal{H})$ is $\mathcal{H}$-Pareto calibrated.*

Thus unrestricted Pareto calibration is a Bayes-level statement, while guarantees for a parameterized neural policy require calibration of the induced class $\mathcal{H}_\Pi$. The latter depends jointly on the surrogate, architecture, parameterization, and any regularization used in training. It cannot in general be inferred from calibration over $\mathcal{H}_{\mathrm{all}}$ alone.

Proof. Write $M_{\mathcal{H}} := \inf_{h \in \mathcal{H}} L_\rho^\phi(h)$. Suppose first that $(\rho, \phi, \mathcal{H})$ is $\mathcal{H}$-Pareto calibrated. Since E_U is finite, $\delta := \min_{e \in E_U} \delta_{\mathcal{H}}(e) > 0$ when E_U is nonempty. If E_U is empty, the conclusion is immediate. If $L_\rho^\phi(h_n) \to M_{\mathcal{H}}$, then for all sufficiently large n,

$$L_\rho^\phi(h_n) < M_{\mathcal{H}} + \delta \le \inf_{\substack{h \in \mathcal{H} \\ m_e(h) \le 0}} L_\rho^\phi(h) \qquad \text{for every } e \in E_U.$$

Consequently $m_e(h_n) > 0$ for every unanimous edge and all sufficiently large n.

Conversely, suppose $\mathcal{H}$-Pareto calibration fails at some $e_0 \in E_U$. Since the constrained feasible set is contained in $\mathcal{H}$, its infimum is at least $M_{\mathcal{H}}$, and failure of a strict gap therefore gives $\inf_{\substack{h \in \mathcal{H} \\ m_{e_0}(h) \le 0}} L_\rho^\phi(h) = M_{\mathcal{H}}$. There is a sequence $\{h_n\} \subseteq \mathcal{H}$ with $m_{e_0}(h_n) \le 0$ and $L_\rho^\phi(h_n) \to M_{\mathcal{H}}$. It is a minimizing sequence over $\mathcal{H}$ but is not asymptotically Pareto optimal. □

G Proofs for Verifying Pareto Calibration

Lemma 5.1. *The Pareto-calibration inequality holds at e_0 if and only if $\Phi'_{e_0}(0^+) < 0$. If the constrained problem at margin zero has a minimizer h_0, then $\Phi'_{e_0}(0^+) \le \frac{1}{2} D(e_0)$.*

Proof. Φ_{e_0} is convex. If $\Phi'_{e_0}(0^+) < 0$ then Φ_{e_0} is strictly decreasing on $(-\infty, 0]$, so $\inf_{s \le 0} \Phi_{e_0} = \Phi_{e_0}(0)$, while $\Phi_{e_0}(s') < \Phi_{e_0}(0)$ for some $s' > 0$. Hence $\inf \Phi_{e_0} < \inf_{s \le 0} \Phi_{e_0}$, the Pareto-calibration inequality. If $\Phi'_{e_0}(0^+) \ge 0$ then Φ_{e_0} is nondecreasing on $[0, \infty)$, so $\inf \Phi_{e_0} = \inf_{s \le 0} \Phi_{e_0}$ and it fails. For the bound, $h_0 + sd$ is feasible for $\Phi_{e_0}(s)$ with equality at $s = 0$, so $\Phi_{e_0}(s) \le L_\rho^\phi(h_0 + sd)$ and $\Phi'_{e_0}(0^+) \le L_\rho^{\phi\prime}(h_0; d) = \frac{1}{2} D(e_0)$ by the definition of $D(e_0)$. □

Lemma 5.2. *For $e_0 = (a, b)$, if $\mathcal{E}(e_0) := \sum_{x \ne a,b} \left[(\rho_{bx} - \rho_{ax})_+ + (\rho_{xa} - \rho_{xb})_+ \right] < 2\kappa \rho_{ab}$, where $G := \sup\{|g| : g \in \partial\phi(\xi),\ \xi \in \mathbb{R}\} \in (0, \infty]$, and $\kappa := |\phi'_+(0)|/G \in [0, 1]$, then the Pareto-calibration inequality holds at e_0. If this holds at every cyclic unanimous edge, then every minimizing sequence of L_ρ^ϕ is asymptotically Pareto optimal. In particular, every minimizer, when one exists, is Pareto optimal.*

Proof. Let h be any score tied at e_0 and put $u_x = h(x) - h(a)$. Differentiability and monotonicity give

$$2L_\rho^{\phi\prime}(h;d) = 2\rho_{ab}\phi'(0) + \sum_{x\neq a,b}\big[(\rho_{ax}-\rho_{bx})\phi'(-u_x) - (\rho_{xa}-\rho_{xb})\phi'(u_x)\big] \tag{15}$$

$$\leq -2\rho_{ab}|\phi'(0)| + G\mathcal{E}(e_0) =: -2c < 0. \tag{16}$$

The bound is uniform over all tied scores. Since $G < \infty$, the continuous monotone derivative ϕ' is uniformly continuous on $\mathbb{R}$. Therefore the first-order remainder in $L_\rho^\phi(h+td)$ is $o(t)$ uniformly over tied h. Taking the infimum over such h gives

$$\Phi_{e_0}(t) \leq \Phi_{e_0}(0) - ct + o(t),$$

so $\Phi'_{e_0}(0^+) \leq -c < 0$. Lemma 5.1 therefore gives calibration at e_0. If the condition holds at every cyclic unanimous edge, Theorem 4.1 gives the sequential conclusion. □

Lemma 5.3. *(i) $\phi_C \in C^1$ and is strictly decreasing. (ii) $\phi_\infty < \phi_C(\xi) < \phi_\infty + C$ for every ξ, with $\phi_C(\xi) \to \phi_\infty + C$ as $\xi \to -\infty$ and $\phi_C(\xi) \to \phi_\infty$ as $\xi \to \infty$. (iii) $\phi_C'(0) = \tau'(\Delta_\phi/C)\phi'(0) < 0$.*

Proof. Set $u(\xi) := (\phi(\xi) - \phi_\infty)/C$. The chain rule gives

$$\phi_C'(\xi) = C\tau'(u(\xi))u'(\xi) = \tau'(u(\xi))\phi'(\xi) < 0.$$

Thus ϕ_C is C^1 and strictly decreasing. Since τ takes values in $[0,1)$, one has $\phi_\infty < \phi_C(\xi) < \phi_\infty + C$ for every finite ξ. The two limits follow from $u(\xi) \to \infty$ as $\xi \to -\infty$ and $u(\xi) \to 0$ as $\xi \to \infty$. Evaluating the derivative at zero gives the displayed formula. □

Lemma 5.4. *For every $C > 0$, the truncated loss ϕ_C is classification-calibrated.*

Proof. Fix $\eta > \frac{1}{2}$. The other case is symmetric. For every $\beta < 0$, strict decrease gives

$$C_\eta^{\phi_C}(\beta) - C_\eta^{\phi_C}(-\beta) = (2\eta-1)\big(\phi_C(\beta) - \phi_C(-\beta)\big) > 0.$$

Moreover, $(C_\eta^{\phi_C})'(0) = (2\eta-1)\phi_C'(0) < 0$, so the unconstrained infimum is strictly below $C_\eta^{\phi_C}(0)$. The wrong-sign conditional risk is continuous on $(-\infty, 0]$, and its limit as $\beta \to -\infty$ is

$$\phi_\infty + \eta C > \phi_\infty + (1-\eta)C = \lim_{\alpha\to\infty} C_\eta^{\phi_C}(\alpha) \geq H^{\phi_C}(\eta).$$

Compactness on a sufficiently large interval together with the strict reflection inequality therefore gives $H^{\phi_C,-}(\eta) > H^{\phi_C}(\eta)$. □

Lemma 5.5. *If ρ is Pareto realizable, then for every $\gamma > 0$ there exists $h_\gamma^U \in \mathcal{H}_{\mathrm{all}}$ with $m_e(h_\gamma^U) \geq \gamma$ for every $e \in E_U$.*

Proof. Since ρ is Pareto realizable, the directed graph G_U is acyclic and admits a topological order. Choose integer ranks $r : \mathcal{X} \to \mathbb{Z}$ with $r(a) \geq r(b) + 1$ whenever $(a,b) \in E_U$. Set $h_\gamma^U := \gamma\, r$. Then for every $e = (a,b) \in E_U$, $m_e(h_\gamma^U) = \gamma\big(r(a) - r(b)\big) \geq \gamma$. □

Theorem 5.1. *Assume ρ is Pareto realizable and $E_U \neq \emptyset$. If $\sum_{e\in E_N}\rho_e C_e < \omega_{\min}^U \min_{e\in E_U}\Delta_{C_e}$ then every minimizing sequence of $L_{\boldsymbol{C}}$ over $\mathcal{H}_{\mathrm{all}}$ is asymptotically Pareto optimal. In particular, every minimizer, if one exists, is Pareto optimal.*

Proof. Write $B := \phi_\infty \sum_e \rho_e$ and $M := \inf_{h\in\mathcal{H}_{\mathrm{all}}} L_{\boldsymbol{C}}(h)$. For any $e_0 \in E_U$ and any score h with $m_{e_0}(h) \leq 0$, strict decrease and the lower bound $\phi_{C_e} > \phi_\infty$ give

$$L_{\boldsymbol{C}}(h) \geq B + \rho_{e_0}\Delta_{C_{e_0}}.$$

By Lemma 5.5, for every $\gamma > 0$ there is a score h_γ^U whose unanimous margins are at least γ. Hence

$$M \leq L_{\boldsymbol{C}}(h_\gamma^U) \leq \sum_{e\in E_U}\rho_e\phi_{C_e}(\gamma) + \sum_{e\in E_N}\rho_e(\phi_\infty + C_e).$$

Letting $\gamma \to \infty$ gives $M \le B + \sum_{e \in E_N} \rho_e C_e$. Thus the gap between the risk of any score violating e_0 and M is at least

$$\rho_{e_0} \Delta_{C_{e_0}} - \sum_{e \in E_N} \rho_e C_e \ge \omega^U_{\min} \min_{e \in E_U} \Delta_{C_e} - \sum_{e \in E_N} \rho_e C_e > 0.$$

The bound is uniform over the finite set E_U, so every minimizing sequence eventually has all unanimous margins positive. Any exact minimizer has the same property. This argument does not require attainment or convexity of the truncated loss. □

H Multi Class Classification Calibration and Pairwise Majority Consistency

Recall the multiclass classification problem defined in Appendix A.

We first verify monotonicity of ρ_{mc}. Fix x and labels a, b with $p_a(x) > p_b(x)$, so $\Delta_{\rho_{\mathrm{mc}}}((x,a),(x,b)) > 0$. For any third label (x,d) at the same feature value,

$$\rho_{\mathrm{mc},(x,d),(x,a)} = P(X = x, Y = d) = \rho_{\mathrm{mc},(x,d),(x,b)}\,,$$

and

$$\rho_{\mathrm{mc},(x,a),(x,d)} = P(X = x, Y = a) \ge P(X = x, Y = b) = \rho_{\mathrm{mc},(x,b),(x,d)}$$

For any alternative (x', d) at a different feature value $x' \neq x$, all masses vanish. Hence ρ_{mc} is monotone at every majority pair.

We use the multiclass calibration definition of Tewari & Bartlett (2007) to compare it with PMC.

For the multi-class pairwise margin loss with binary surrogate ϕ, define $\Psi : \mathbb{R}^K \to \mathbb{R}^K_+$ by

$$\Psi_y(f) \;:=\; \sum_{b \neq y} \phi(f_y - f_b), \qquad y \in [K],\ f \in \mathbb{R}^K,$$

and let $\mathcal{R}_\phi := \{\Psi(f) : f \in \mathbb{R}^K\}$, $\mathcal{S}_\phi := \mathbf{conv}(\mathcal{R}_\phi) \subseteq \mathbb{R}^K_+$. The conditional surrogate risk at x is

$$F_x(f) \;=\; \sum_a p_a(x)\, \Psi_a(f) \;=\; \langle p(x), \Psi(f) \rangle, \qquad \inf_{f \in \mathbb{R}^K} F_x(f) = \inf_{z \in \mathcal{S}_\phi} \langle p(x), z \rangle,$$

where the second equality uses that the infimum of a linear functional over $\mathcal{R}_\phi$ equals that over its convex hull. The loss treats labels symmetrically. We regard a loss vector z as making an incorrect prediction whenever at least one label minimizing z_y is not Bayes optimal.

Definition H.1 (Multiclass calibration Tewari & Bartlett (2007))**.** *For $p \in \Delta_K$, the loss Ψ is classification calibrated at p if*

$$\inf_{\substack{z \in \mathcal{S}_\phi:\ \exists y \in \mathrm{argmin}_j\, z_j \\ p_y < \max_j p_j}} \langle p, z \rangle > \inf_{z \in \mathcal{S}_\phi} \langle p, z \rangle. \tag{17}$$

It is classification calibrated if this holds for every $p \in \Delta_K$.

Theorem H.1. *Let $\Psi_y(f) = \sum_{b \neq y} \phi(f_y - f_b)$.*

1. *Assume the feature support is finite, $p_y(x) > 0$ for every supported x and class y, and $\phi : \mathbb{R} \to \mathbb{R}_+$ is continuous and strictly decreasing with $\phi(t) \to \infty$ as $t \to -\infty$. If every minimizer of $L^\phi_{\rho_{\mathrm{mc}}}$ over $\mathcal{H}_{\mathrm{all}}$ satisfies PMC, then Ψ is classification calibrated at every supported $p(x)$.*

2. *There are a loss ϕ and $p \in \Delta_K$ for which Ψ is classification calibrated at p, while every conditional-risk minimizer violates PMC at p.*

Proof. For the first claim, fix a supported feature value x and write $p = p(x)$ and $F(f) = \langle p, \Psi(f) \rangle$. Translation of all coordinates of f leaves F unchanged. Since every $p_a > 0$ and $\phi(t) \to \infty$ as $t \to -\infty$, every sublevel set of F is bounded after normalizing $\max_y f_y = 0$. Bounded risk bounds each $\phi(f_a - f_b)$ and hence every score difference from below, while interchanging a, b

bounds it above. Thus F attains its minimum F^*, and its normalized minimizer set is compact. Because the feature support is finite, every conditional minimizer extends to a global minimizer by selecting minimizers at the other feature values. The hypothesis therefore gives strict PMC at every minimizer of F. For every Bayes label i and non-Bayes label y, strict PMC implies $f_i > f_y$, hence $\Psi_i(f) < \Psi_y(f)$. Compactness gives a uniform positive gap δ between these loss coordinates over all minimizers. By continuity and compactness of a low-risk sublevel set, there is an $\eta > 0$ such that $F(f) \le F^* + \eta$ implies $\Psi_y(f) - \Psi_i(f) \ge \delta/2$ for every such i, y.

Now suppose calibration fails. Choose bad vectors $z_n \in \mathbf{conv}\{\Psi(f) : f \in \mathbb{R}^K\}$ with $\langle p, z_n\rangle \to F^*$. Write each z_n as a finite convex combination of $\Psi(f)$, and split the terms into $F(f) \le F^* + \eta$ and $F(f) > F^* + \eta$. The total weight of the latter is at most $(\langle p, z_n\rangle - F^*)/\eta \to 0$. Since all loss coordinates are nonnegative and $p_{\min} > 0$, their contribution to each coordinate is at most $[\langle p, z_n\rangle - F^* + F^* \times \text{their total weight}]/p_{\min} \to 0$. The low-risk terms retain the gap $\delta/2$ between every non-Bayes coordinate and every Bayes coordinate. Hence, for large n, no non-Bayes label minimizes z_n, a contradiction.

For the second claim, take $\phi(t) = (1-t)_+$, $K = 3$, and $p = (1/2, 3/10, 1/5)$. For any probability vector p' and nonnegative numbers $q_{ab} \le p'_a$ satisfying flow balance $\sum_{b\neq a} q_{ab} = \sum_{b\neq a} q_{ba}$, the hinge inequality gives

$$\langle p', \Psi(f)\rangle \ge \sum_{a\neq b} q_{ab}(1 - f_a + f_b) = \sum_{a\neq b} q_{ab}.$$

The following table lists feasible certificates $q_{12}, q_{13}, q_{21}, q_{23}, q_{31}, q_{32}$ together with the resulting lower bounds.

p'	q_{12}	q_{13}	q_{21}	q_{23}	q_{31}	q_{32}	$\sum q_{ab}$
$(1/2, 3/10, 1/5)$	$2/5$	$1/10$	$3/10$	$3/10$	$1/5$	$1/5$	$3/2$
$(7/15, 1/3, 1/5)$	$7/15$	$1/15$	$1/3$	$1/3$	$1/5$	$1/5$	$8/5$
$(11/30, 3/10, 1/3)$	$4/15$	$11/30$	$3/10$	$3/10$	$1/3$	$1/3$	$19/10$

The first bound is attained at $f^* = (1, 0, 0)$. In its certificate, q_{12} and q_{13} are strictly between zero and p_1, so equality forces $1 - f_1 + f_2 = 1 - f_1 + f_3 = 0$. Thus f^* is the unique minimizer modulo translation. It ties classes 2 and 3 despite $p_2 > p_3$, violating PMC.

The second and third bounds are attained at $f = (0, 0, -1)$ and $f = (0, -1, 0)$, respectively. For every $z \in \mathcal{S}_\phi$, the second certificate gives $\langle p, z\rangle + \frac{1}{30}(z_2 - z_1) \ge 8/5$, and the third gives $\langle p, z\rangle + \frac{2}{15}(z_3 - z_1) \ge 19/10$. If $z_2 \le z_1$ or $z_3 \le z_1$, respectively, then $\langle p, z\rangle$ is at least the corresponding bound. Every loss vector with an incorrect minimizing label satisfies one of these inequalities, while the unrestricted infimum is $3/2$. This proves calibration at p even though PMC fails at its unique score minimizer. □

I Data Reweighting

I.1 Data reweighting

Split the surrogate risk into its unanimous and non-unanimous parts,

$$L_\rho^\phi(h) = \sum_{e\in E_U} \rho_e\, \phi(m_e(h)) + \sum_{e\in E_N} \rho_e\, \phi(m_e(h)) =: L_U(h) + R(h).$$

A reweighting of the unanimous data is a vector $\lambda = (\lambda_e)_{e\in E_U}$ with $\lambda_e \ge 0$. It replaces the mass ρ_e on each unanimous edge, giving

$$L_\lambda(h) = \sum_{e\in E_U} \lambda_e\, \phi(m_e(h)) + \sum_{e\in E_N} \rho_e\, \phi(m_e(h)).$$

If the unanimous edge has weight Λ instead of 2 and all other weights remain fixed, the minimizing margin becomes $h(a) - h(b) = (2\Lambda - 23)/(2\Lambda + 25/2)$. It is positive once $\Lambda > 23/2$. This calculation illustrates reweighting for the squared loss. The theorem below separately covers strictly decreasing losses.

Throughout this subsection we assume $\phi : \mathbb{R} \to \mathbb{R}$ is convex, strictly decreasing, and bounded below with $\phi_\infty := \inf_{\mathbb{R}} \phi > -\infty$, that ρ is Pareto realizable, and that $E_U \neq \emptyset$. No attainment of L_Λ is assumed. Write $n_U := |E_U|$.

Theorem I.1 (Static reweighting). *Let $R_{\min} := \inf_{h \in \mathcal{H}_{\text{all}}} R(h)$ and choose $\gamma > 0$ such that*

$$d(\gamma) := [\phi(0) - \phi(\gamma)] - (n_U - 1)[\phi(\gamma) - \phi_\infty] > 0.$$

Such a γ exists because $d(\gamma) \to \phi(0) - \phi_\infty > 0$. Set $\Lambda^* := [R(h_\gamma^U) - R_{\min}]/d(\gamma)$. For every $\Lambda > \Lambda^*$, the uniformly reweighted risk L_Λ has a positive gap between every PO-violating score and its infimum. Consequently every minimizing sequence is asymptotically Pareto optimal, and every minimizer, if one exists, is Pareto optimal.

Proof. Write $S(h) := \sum_{e \in E_U} \phi(m_e(h))$, so $L_\Lambda = R + \Lambda S$. Lemma 5.5 gives a score h_γ^U with $S(h_\gamma^U) \le n_U \phi(\gamma)$, hence

$$\inf_h L_\Lambda(h) \le R(h_\gamma^U) + \Lambda n_U \phi(\gamma).$$

If h violates any unanimous edge, then $S(h) \ge \phi(0) + (n_U - 1)\phi_\infty$ and $R(h) \ge R_{\min}$. Therefore

$$L_\Lambda(h) - \inf_{h'} L_\Lambda(h') \ge \Lambda d(\gamma) - [R(h_\gamma^U) - R_{\min}] = d(\gamma)(\Lambda - \Lambda^*) > 0.$$

The bound is uniform over all violating scores, so every minimizing sequence eventually has all unanimous margins positive. □

Since Λ^* depends on ϕ and the comparison structure, an approximate optimization procedure can increase Λ until its returned score has positive unanimous margins.

Lemma I.1 (Adaptive reweighting). *Fix $c > 1$, $\Lambda_0 > 0$, and an optimization tolerance $\epsilon \ge 0$. At step $t = 0, 1, \ldots$, choose a score $h^{(t)}$ satisfying $L_{\Lambda_t}(h^{(t)}) \le \inf_h L_{\Lambda_t}(h) + \epsilon$, where $\Lambda_t = c^t \Lambda_0$. Stop when $h^{(t)}$ is Pareto optimal. Otherwise, increase the weight. If $\epsilon = 0$, assume the selected exact minimizer exists. Set $T_\epsilon := \Lambda^* + \epsilon/d(\gamma)$. The stopping index obeys*

$$t_{\text{stop}} \le \begin{cases} 0, & T_\epsilon = 0, \\ \max\{0, \lfloor \log_c(T_\epsilon/\Lambda_0) \rfloor + 1\}, & T_\epsilon > 0, \end{cases}$$

so at most $t_{\text{stop}} + 1$ scores are computed. The returned score is Pareto optimal.

Proof. The gap in Theorem I.1 is $d(\gamma)(\Lambda_t - \Lambda^*)$. Whenever $\Lambda_t > T_\epsilon$, this exceeds ϵ. Therefore no ϵ-optimal score can violate a unanimous edge. If $T_\epsilon = 0$, step 0 already suffices. Otherwise the first integer $t \ge 0$ with $c^t \Lambda_0 > T_\epsilon$ is $\max\{0, \lfloor \log_c(T_\epsilon/\Lambda_0) \rfloor + 1\}$, including when the ratio is an exact power of c. The stopping test succeeds by that step. □

J EXPERIMENT

Implementation details. Every run uses LoRA with rank 16, $\alpha = 32$, dropout 0.05, and target modules $q, k, v, o, \text{gate}, \text{up}, \text{down}$. The optimizer is AdamW at learning rate 10^{-4} with no weight decay, a 10% linear warmup followed by cosine decay, gradient clipping 1.0, global batch size 128 pairs, and three epochs (147 updates per epoch, 441 updates in total). Checkpoints and reweighting occur twice per epoch. All models are trained with DeepSpeed ZeRO stage 2 on one L40S GPU.

Each evaluation is done with the model's chat template. Response log probabilities are sums over response tokens only. Sequences are truncated to $1{,}024$ tokens with at most 512 prompt tokens, keeping the final prompt tokens and the response prefix and forcing an EOS token. Reference log probabilities are precomputed once from the frozen initialization and reused by every method.

Compute Each training job used one NVIDIA L40S GPU, eight CPU cores, and 64 GB of host memory at 0.5B or 96 GB at 3B and 7B, with DeepSpeed ZeRO stage 2.