When Pretraining Votes Twice: Social Choice Mechanics of Pooled Human Cognition in Language Models

Nyx Redondo[1,2]

[1]Ronin Institute (RIIS 2.0) [2]IamI.Earth Foundation, Stockholm, Sweden
nyx.redondo@ronininstitute.org · nyx@iami.earth · ORCID: 0009-0004-2876-0025
Preprint v1.0, November 2026 · doi:10.5281/zenodo.24571109

Abstract

Random initial weights carry zero world bits; every fact, rule, and preference enters from human text and design, reorganized by computation. Modeling the three-stage language-model pipeline (pretraining, preference fitting, and regularized alignment) as coupled social choice, we prove pretraining votes twice: as *strength of schedule* (choosing which answers are compared) and *home-field pull* (anchoring the Kullback-Leibler leash). On a hand-checkable 45-rater example where 77.8% prefer a over b_1, either handicap alone cuts a's 4:1 lead roughly $3\times$ (1.31:1 and 1.30:1, leaving a ahead), yet both multiply to flip the winner to b_1 by 2.36:1. More strongly, a corpus niche under 0.4% overturns a 60% majority, and any supermajority can collapse near zero. We prove Stage 1 pooling synthesizes unwritten answers, extend the mechanics to DPO, GRPO, and pooling order, and show finite-leash clone invariance requires declared class-standing weights via a two-level hierarchical policy.

1 Introduction: Two Puzzles, One Voting Machine

Two puzzles fuel the belief that post-trained language models hold goals of their own: a model can write a right answer found in zero training documents, yet after Reinforcement Learning from Human Feedback (RLHF) overturn an answer favored by nearly four in five raters.

Both dissolve when we view the pipeline as a voting machine (Figure 1). Starting from zero world bits, Stage 1 pools human text under cross-entropy compression, which can peak on an unwritten consensus. Stages 2 and 3 fit pairwise ratings and tilt the pretrained pool m_x under a Kullback-Leibler (KL) leash. Because standard pipelines reuse m_x in both stages, *pretraining votes twice*.

A 10-second preview shows both votes compounding. Suppose 77.8% of raters prefer a over b_1, giving a a 4:1 lead under balanced comparisons and a uniform leash, alongside a weak decoy b_2 that b_1 crushes more sharply than a does. **Vote 1 (strength of schedule)** draws pairs from m_x: beating b_2 pads b_1's rating, cutting a's lead roughly $3\times$ to 1.31:1 (a still wins). **Vote 2 (home-field pull)** anchors the KL leash to m_x, again cutting a's lead roughly $3\times$ to 1.30:1 (a still wins). Coupled, both handicaps multiply $> 9\times$, flipping the winner to b_1 by 2.36:1 on identical ratings (Table 3).

Prior work and contributions. Single stages are known: cross-entropy targets linear opinion pooling (Abbas 2009), product-sampled Bradley-Terry (BT) scores track menu-weighted Borda counts (Siththaranjan, Laidlaw, and Hadfield-Menell 2024), and KL alignment yields logarithmic or rational pooling (Dietrich and List 2016; Korbak, Perez, and Buckley 2022; Vojnovic and Yun 2025). Recent work examines alignment paradoxes and clones (Conitzer et al. 2024; Ge et al. 2024; Procaccia, Schiffer, and Zhang 2025), maximal lotteries (Fishburn 1984; Brandl, Brandt, and Seedig 2016), and models as judgment aggregators (Betz and Richardson 2022) or pooled human cognition (Redondo 2026).

Composing all three stages yields four results: (1) an exact decoy cliff τ and global niche formula (Theorem 2); (2) a $4{:}1 \to 2.36{:}1$ double-voting witness and supermajority-collapse theorem below τ (Table 3, Theorem 3); (3) a 4Δ BT-Borda bound and equal-match Condorcet flip (Table 2); and (4) proof that finite-leash clone invariance fails under global homotopy and requires declared standing ρ, realized by a stable hierarchical policy (Theorem 4, Appendix C).

2 Zero-Start Accounting vs. Computation

Separating *where world facts enter* from *how computation makes patterns readable* resolves a common confusion. Draw independent fair bits $S \in \{0,1\}$ and $X_1 \in \{\pm 1\}$, and set $X_2 = (2S-1)X_1$ ($S = 1$ iff both coins match). In Shannon mutual information $I(\cdot;\cdot)$ (0 means unrelated), each coin alone reveals 0 bits about S, and every linear rule guesses $50/50$ (0 linearly usable bits). Yet $Z_1 = \max(0, X_1 + X_2)$ and $Z_2 = \max(0, -X_1 - X_2)$ obey $Z_1 + Z_2 = 2S$, unlocking the full 1 linear bit without importing any bit beyond $X = (X_1, X_2)$.

Proposition 1 (Source Closure and Usable Extraction)**.** *Let H, I be bit entropy and mutual information for discrete N-fact state $S = S^{(N)}$ ($H(S) < \infty$), human dataset $\mathcal{D}$, human-written code $\mathcal{T}$ of prefix-free length $\ell(\mathcal{T})$, independent seed $\xi \perp (\mathcal{D}, \mathcal{T}, S)$, initial weights $\Theta_0 = g_0(\xi)$, and closed deterministic training $\Theta_T = F(\Theta_0, \mathcal{D}, \mathcal{T}, \xi)$ (general $\mathcal{E}$ in Appendix A). **(a) Source closure:** $I(\Theta_0; S) = 0$ and*

$$I(\Theta_T; S^{(N)}) \le I(\mathcal{D}; S^{(N)}) + \mathbb{E}[\ell(\mathcal{T})] \tag{1}$$

*(Cover and Thomas 2006; Grünwald 2007), so $\mathbb{E}[\ell(\mathcal{T})] = o(N)$ caps per-fact bits at $I(\mathcal{D}; S^{(N)})/N + o(1)$. **(b) Usable***

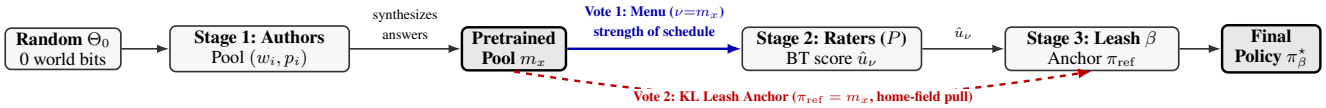

Figure 1: Coupled pipeline. From $I(\Theta_0; S) = 0$, Stage 1 pools authors (w_i, p_i) into m_x; alignment lets m_x vote twice via Stage 2's menu (ν=m_x) and Stage 3's KL anchor (π_{ref}=m_x; smaller β loosens the leash).

Checklist Answer (each)	Raw $\bar{p}_w$	Factored $q^\star$
$(\mathbf{Y},\mathbf{Y},\mathbf{Y})$ **(unwritten)**	$\mathbf{0.0\%}$	$\mathbf{8/27\ (29.6\%)}$
$(Y,Y,N), (Y,N,Y), (N,Y,Y)$	$1/3\ (33.3\%)$	$4/27\ (14.8\%)$
$(Y,N,N), (N,Y,N), (N,N,Y)$	0.0%	$2/27\ (7.4\%)$
(N,N,N)	0.0%	$1/27\ (3.7\%)$

Table 1: Checklist synthesis ($2/3$ Y per slot): factored $q^\star$ lifts unwritten (Y,Y,Y) to unique mode $(8/27)$.

extraction: *For $S - (X, \mathcal{D}, \mathcal{T}) - Z$ and predictive family $\mathcal{V}$ with optional ignorance, usable $\mathcal{V}$-information (Xu et al. 2020) obeys $I_{\mathcal{V}}(Z \to S) \le I(Z; S) \le I(X, \mathcal{D}, \mathcal{T}; S)$, yet on the two-coin puzzle $Z = (Z_1, Z_2)$ lifts logistic-linear $I_{\mathcal{V}_{\text{lin}}}$ from 0 to 1 bit (Appendix A).*

Deterministic code imports zero outside facts, yet nonlinear circuits turn buried parity into a linear threshold. The same holds for math and code verifiers: across tasks, answer S has $H(S) > 0$ yet $H(S \mid \mathcal{D}, \mathcal{T}) = 0$, so $I(\mathcal{D}, \mathcal{T}; S) = H(S)$; search and verifier RL turn rule-fixed structure into usable $\mathcal{V}$-information.

3 Stage 1: Crowd Pooling, Synthesis, and Routing

A crowd beats its average member by exactly how much its members disagree (ln throughout). Three examples illustrate pooling, synthesis, and routing.

1. Crowd averaging ($10 = 1{+}9$). For true weight 6, equal-share guessers saying 2 (squared error 16) and 8 (4) average error 10, whereas their midpoint 5 has error 1, beating *both* by their variance $\frac{1}{2}(2{-}5)^2{+}\frac{1}{2}(8{-}5)^2 = 9$ (Krogh and Vedelsby 1995; Hong and Page 2004). At Plymouth, Galton (1907) found 787 ox guesses hit median 1,207 vs. true 1,198 lb (0.8% error). Fitting one unit-variance Gaussian peaks at unwritten 5.

2. Unwritten checklist consensus ($8/27$ vs. $4/27$). Lookup tables memorize raw cards (33.3% each); compressed models pool shared parts. In Table 1, three equal-share experts fill a truth-(Y,Y,Y) checklist with $2/3$ Y per slot. Factored fitting $\prod_{k=1}^{3} q_k(y_k)$ lifts unwritten (Y,Y,Y) to unique mode $(2/3)^3 = 8/27$, double any written card. Slotwise pooling cuts both ways (Dietrich and List 2016; Betz and Richardson 2022): independent slots synthesize truth, whereas a hidden rule ("at most two Ys") yields a discursive-dilemma output endorsed by zero authors.

3. Prompt routing (90% specialist filtering). If specialists and novices split global text 50/50, yet specialists write on prompt x nine times as often (0.9 vs. 0.1), Bayes' rule lifts prompt-x specialist share to 90%.

Theorem 1 (Stage 1 Projection, Synthesis, and Routing)**.** *On $(\mathcal{Y}, \mu)$ (counting or Lebesgue on $\mathbb{R}$), let authors $i = 1, \dots, K$ have μ-densities p_i, shares $w \in \operatorname{int} \Delta_K$, and pool $\bar{p}_w := \sum_i w_i p_i$.* ***(a) KL projection split:*** *Cross-entropy targets $\bar{p}_w$ (Abbas 2009). For any minimal regular exponential family $\mathcal{Q}$ with $\bar{\mu}_w := \mathbb{E}_{\bar{p}_w}[\phi(Y)]$ in its interior mean range, $q^\star = \arg\min_{q \in \mathcal{Q}} \mathrm{KL}(\bar{p}_w \| q)$ is unique, matches $\mathbb{E}_{q^\star}[\phi(Y)] = \bar{\mu}_w$, and for all $q \in \mathcal{Q}$ with finite $\mathrm{KL}(p_i \| q)$ (Banerjee et al. 2005),*

$$\sum_i w_i \,\mathrm{KL}(p_i \| q) = \underbrace{\sum_i w_i \,\mathrm{KL}(p_i \| \bar{p}_w)}_{\textit{disagreement}} + \underbrace{\mathrm{KL}(\bar{p}_w \| q^\star)}_{\textit{synthesis gap}} + \underbrace{\mathrm{KL}(q^\star \| q)}_{\textit{crowd error}}. \tag{2}$$

(b) Support gate and routing: *Set $m_x := q^\star(\cdot \mid x)$ ($\bar{p}_w(\cdot \mid x)$ if unconstrained). Any finite-KL Stage 3 policy anchored at m_x obeys $\operatorname{supp}(\pi^\star_\beta(\cdot \mid x)) \subseteq \operatorname{supp}(m_x)$, so compression unlocks unwritten answers. When $\bar{p}_w(x) > 0$, prompt x reweights authors to $\omega_i(x) := w_i p_i(x)/\bar{p}_w(x)$ (stability: Prop. A.1).*

Doubling (2) on the Gaussians ($2, 8$ vs. 6) recovers $10 = (1.38{+}7.62){+}1$, while Table 1 splits $\ln 8 = \ln 3{+}\ln(9/4){+}\ln(32/27)$ against uniform $q_{\text{unif}} \equiv 1/8$ (Appendix A).

4 Stages 2 and 3: How Pretraining Votes Twice

Fix prompt x, finite slate A ($|A| \ge 2$), and pairwise human win rates $P(y, y') \in [0, 1]$ ($P(y, y') + P(y', y) = 1$, $P(y, y) = 1/2$).

4.1 Vote 1: Strength of Schedule and Decoy Cliffs

Vote 1 is strength of schedule: beating weak decoys pads ratings. Stage 2 BT fitting (Zermelo 1929; Bradley and Terry 1952; Ford 1957) assigns odds $r_y := e^{\hat{u}(y)}$ (zero-sum $\hat{u}$) matching predicted wins $\sum_{y' \ne y} N_{yy'} \frac{r_y}{r_y + r_{y'}}$ to human wins $\sum_{y' \ne y} N_{yy'} P(y, y')$ on symmetric counts $N_{yy'} = N_{y'y} \ge 0$ ($\sigma(z) := (1{+}e^{-z})^{-1}$ has peak slope $1/4$). If P already fits a BT scale, the menu drops out; Vote 1 activates when diverse raters depart from one scale.

Under product sampling ($N_{yy'} = \nu(y)\nu(y')$), BT tracks menu-weighted Borda score $B_\nu(y) := \sum_{y'} \nu(y') P(y, y')$ (Siththaranjan, Laidlaw, and Hadfield-Menell 2024), stretching any gap Δ to $\ge 4\Delta$. Even with equal match counts, a curated schedule can invert a Condorcet winner (Table 2):

Ans.	Win rates $P(y,\cdot)$	Unif.	Cur.
a (CW)	b,d:.70; c:.55	**.6125**	12.40
b (BT)	a:.30; c:.60; d:.85	.5625	**17.90**
c	a:.45; b,d:.40	.4375	9.80
d	a:.30; b:.15; c:.60	.3875	3.90

Table 2: Curated flip on 100 transitive raters (30 $b\succ d\succ c\succ a$, 30 $a\succ b\succ d\succ c$, 25 $a\succ c\succ b\succ d$, 15 $c\succ a\succ d\succ b$; CW: Condorcet). Pairing $(a,c),(b,d)$ 20× and others 1× (22 each) makes BT pick b.

pairing b with weak d (20×) and a with tough c (20×) lifts b's wins to 17.90 vs. 12.40, ranking b first (Ge et al. 2024).

Theorem 2 (Vote 1: Schedule Power and Decoy Cliff)**.** *Let symmetric $N_{yy'}=N_{y'y}\geq 0$ have strongly connected positive-win graph $G_+=\{(y,y'):N_{yy'}P(y,y')>0\}$, so maximizing $\mathcal{L}_{\mathrm{BT}}(u)=\sum_{y,y'}N_{yy'}P(y,y')\ln\sigma(u(y)-u(y'))$ gives unique zero-sum $\hat{u}_N$. **(a) Invariance, 4Δ gap, and flips:** (i) If $P(y,y')=\sigma(u^\star(y)-u^\star(y'))$ (zero-sum $u^\star$), then $\hat{u}_N=u^\star$ for all such N. (ii) For full-support $N_{yy'}=\nu(y)\nu(y')$, $\hat{u}_\nu$ increases strictly with B_ν (Siththaranjan, Laidlaw, and Hadfield-Menell 2024) and $|\hat{u}_\nu(y)-\hat{u}_\nu(y')|\geq 4|B_\nu(y)-B_\nu(y')|$. (iii) Equal-match schedules can invert a transitive Condorcet and uniform-Borda winner (Table 2). **(b) Decoy cliff τ and niche leverage ($\beta\to 0^+$):** On $A=\{a,b_1,b_2\}$ with $P(a,b_k)=\alpha_k\in(1/2,1)$ (k=1, 2), $P(b_1,b_2)=\gamma\in[1/2,1]$, and $\nu=\pi_{\mathrm{ref}}=m_x\in\operatorname{int}\Delta(A)$: if $\gamma\leq\alpha_2$ (including clones $\gamma=1/2$) then $\pi^\star_\beta(a)\to 1$; if $\gamma>\alpha_2$ then $\pi^\star_\beta(b_1)\to 1$ iff*

$$m_x(b_2)>\tau:=\frac{\alpha_1-1/2}{(\alpha_1-1/2)+(\gamma-\alpha_2)}. \tag{3}$$

If $m_x=\bar{p}_w(\cdot\mid x)$ and only author j writes b_2 at x with rate $e_j:=p_j(b_2\mid x)\in(\tau,1]$ and topic concentration $\Lambda_j(x):=p_j(x)/\bar{p}_{-j}(x)>0$ (non-j mean $\bar{p}_{-j}$), then (3) *holds iff $w_j>\tau/[\tau+(e_j-\tau)\Lambda_j(x)]$ (Appendix B).*

Each slice of b_2 adds γ wins to b_1 vs. α_2 to a, crossing at τ. **Headline threshold:** if 60% prefer a over b_1,b_2 while b_1 beats b_2 by 90%, $\tau=0.10/(0.10+0.30)=25\%$. At 100× topic concentration ($e_j=1$), just 0.332% (1/301) of global text flips BT to b_1 (0.249%, 1/401, at $\gamma=100\%$, $\tau=20\%$).

4.2 Vote 2 and Compounding: Home-Field Pull Multiplies

Vote 2 is home-field pull: Stage 3 outputs $\pi^\star_\beta(y)\propto\pi_{\mathrm{ref}}(y)e^{\hat{u}_\nu(y)/\beta}$ (Korbak, Perez, and Buckley 2022; Dietrich and List 2016), tugging probabilities toward pretraining shares. At leash $\beta=1/2$, this multiplies anchor weight $\pi_{\mathrm{ref}}(y)$ by squared BT odds r_y^2.

Table 3 uses 45 transitive raters (31 $a\succ b_1\succ b_2$, 6 $b_1\succ b_2\succ a$, 4 $b_1\succ a\succ b_2$, 4 $b_2\succ a\succ b_1$; zero Pareto dominance). Here a beats b_1,b_2 by $35/45=7/9$ (77.8%) while b_1 beats b_2 by 41/45 (91.1%), so $\tau=25/37\approx 67.6\%$. Set $m_x=(13,40,88)/141\approx(9.2\%,28.4\%,62.4\%)$, safely *below* τ (1.63% niche at 100× concentration). Decoupled (Row 1), squared odds $(64,16,1)$ give a a 4.00:1 **lead** (79.0% vs. 19.8%). **Vote 1 alone** (Row 2) squeezes odds to $(8,7,1)$ (Tideman 1987; Procaccia, Schiffer, and Zhang 2025), cutting a's lead 3.06× to $64/49\approx\mathbf{1.31}$:1 ($a$ **still wins**). **Vote 2 alone** (Row 3) tilts the anchor 40:13, cutting a's lead 3.08× to **1.30**:1 (a **still wins**). Coupled (Row 4) multiplies both (9.42×, factor $26/245$) to $104/245\approx 0.424$:1: b_1 **wins by** 2.36:1 (68.1% vs. 28.9%), while **first-choice satisfaction** (expected rater share getting its top pick) drops from 58.9% to 35.3% (Appendix B).

Theorem 3 (Double-Voting Compounding and Collapse)**.** *On $A=\{a,b_1,b_2\}$, fix $1/2<\alpha<\gamma<1$ with $P(a,b_1)=P(a,b_2)=\alpha$, $P(b_1,b_2)=\gamma$, and $\tau=(\alpha-1/2)/(\gamma-1/2)>1/3$ (automatic for $\alpha\geq 2/3$). Set $\Delta_\nu:=\hat{u}_\nu(a)-\hat{u}_\nu(b_1)$, $\kappa:=\ln(m_x(b_1)/m_x(a))$, and $\pi^\star_\beta(y)\propto\pi_{\mathrm{ref}}(y)e^{\hat{u}_\nu(y)/\beta}$. **(a) Compounding band:** For $m_x\in\operatorname{int}\Delta(A)$ with $m_x(b_1)>m_x(a)$, $m_x(b_2)<\tau$, and $\Delta_{m_x}<\Delta_{\mathrm{unif}}$, coupled RLHF makes b_1 unique mode while both single-vote controls keep a unique mode iff $\beta\in(\Delta_{m_x}/\kappa,\min(\Delta_{\mathrm{unif}}/\kappa,\beta_{\mathrm{mode}}))$ (β_{mode} in Appendix B), nonempty iff $\Delta_{m_x}/\kappa<\beta_{\mathrm{mode}}$ (guaranteed as $m_x(b_2)\uparrow\tau$; Table 3 gives $(0.119,0.617)$, $\beta_{\mathrm{mode}}\approx 1.087$). **(b) Supermajority collapse:** For every such (α,γ) and $\eta\in(0,1/4)$, there exist $m_x\in\operatorname{int}\Delta(A)$ ($m_x(b_2)<\tau$) and $\beta>0$ where each single-vote control gives a mass $>1-\eta$, while coupled RLHF drives $\pi^\star_\beta(a)<\eta$ and $\pi^\star_\beta(b_1)>1-\eta$ (Appendix B).*

Pushing $m_x(b_2)$ near τ shrinks Δ_{m_x} toward zero while Δ_{unif} stays wide, so a moderate anchor tilt flips any supermajority only when both votes stack.

4.3 DPO, GRPO, and Pooling-Order Reversals

Direct Preference Optimization (DPO) inherits this Gibbs limit (Rafailov et al. 2023), and RLAIF iterates it on model-judged pairs. In Group Relative Policy Optimization (GRPO), the stop-gradient reverse-KL penalty gives the rational fixed point $\pi^\star_{\mathrm{GRPO}}(y)=\beta m_x(y)/(\lambda+\beta-\tilde{R}(y))$ for fixed group-standardized advantage $\tilde{R}$ ($\lambda>\max_y\tilde{R}(y)-\beta$, Vojnovic and Yun 2025, Appendix C): m_x anchors the numerator (Vote 2), while m_x rollouts set $\tilde{R}$'s baseline.

Moreover, *when* diverse rater groups are pooled flips the winner (Chakraborty et al. 2024; Abbas 2009; Dietrich and List 2016). Suppose three equal groups align $\{a,b\}$ at $\beta=1$ (uniform anchor) and give a probabilities 96%, 25%, **and** 25% (odds 24:1, 1:3, 1:3). *Post-alignment linear pooling* averages to $(96+25+25)/3=48.7\%$ (73/150), so b **wins**; *pre-alignment log-odds pooling* takes the geometric mean of odds, $(8/3)^{1/3}\approx 1.39$:1 (58.1%), so a **wins** for every $\beta>0$.

4.4 Clone Inflation and Three-Tier Decoupling

Splitting one stance into paraphrases swells its raw pretraining share, a social-choice failure called **clone inflation** (Tideman 1987; Procaccia, Schiffer, and Zhang 2025). Table 4 and Theorem 4 present a three-tier fix.

Theorem 4 (Three-Tier Decoupling and Standing Necessity)**.** *Fix $m_x\in\operatorname{int}\Delta(A)$ and clone partition $\mathcal{C}$ of A ($|\mathcal{C}|\geq 2$): $P(y,y')=P_{\mathcal{C}}(C,C')$ for all $C\neq C'$, $y\in C,y'\in C'$. **(i) Tier 1 (Balanced menu):** Any m_x-independent full-support ν_{bal} with strongly connected*

Row	Pipeline Wiring	Menu ν	Anchor π_{ref}	BT r	$\pi_{\text{ref}}\, r^2$ ($\beta{=}\frac{1}{2}$)	Policy (a, b_1, b_2)	Winner (1st-choice)
1	Decoupled (control)	Unif.	Unif.	$(8, 4, 1)$	$(64, 16, 1)$	$(\mathbf{79.0}\%, 19.8\%, 1.2\%)$	**a** 4.00:1 (58.9%)
2	Vote 1 only (schedule)	m_x	Unif.	$(8, 7, 1)$	$(64, 49, 1)$	$(\mathbf{56.1}\%, 43.0\%, 0.9\%)$	**a** 1.31:1 (48.3%)
3	Vote 2 only (leash)	Unif.	m_x	$(8, 4, 1)$	$(832, 640, 88)$	$(\mathbf{53.3}\%, 41.0\%, 5.6\%)$	**a** 1.30:1 (46.4%)
4	**Coupled (both votes)**	$\mathbf{m_x}$	$\mathbf{m_x}$	$(\mathbf{8}, \mathbf{7}, \mathbf{1})$	$(\mathbf{832}, \mathbf{1960}, \mathbf{88})$	$(\mathbf{28.9}\%, \mathbf{68.1}\%, \mathbf{3.1}\%)$	$\mathbf{b_1}$ **2.36:1 (35.3%)**

Table 3: Double-voting witness ($\beta{=}1/2$, 45 raters: $a \succ b_1, b_2$ 77.8%; $b_1 \succ b_2$ 91.1%; $\tau{=}67.6\%$; $m_x{=}(13, 40, 88)/141$, $62.4\% {<} \tau$). Stacking $3.06\times$ and $3.08\times$ handicaps ($9.42\times$) flips 4:1 to 2.36:1 (first-choice down > 11 pts).

Architecture	Vote 1	Vote 2	Needs
Coupled (m_x, m_x)	Active	Active	Raw m_x
Tier 1: Bal. menu	Off	Finite β	Fixed ν_{bal}
Tier 2: Max. lott.	Off ($\beta{\to}0$)	Unique $\mathcal{M}$	Pairwise P
Tier 3: Hier. quot.	Off	Set by ρ	Classes $\mathcal{C}, \rho$

Table 4: Three-tier decoupling (Thm. 4): Tier 2 needs $\beta \to 0^+$ and unique $\mathcal{M}(P_{\mathcal{C}})$; finite β needs standing ρ.

$G_+(\nu_{\text{bal}})$ makes $\hat{u}_{\nu_{\text{bal}}}$ independent of m_x, cutting Vote 1 while keeping Vote 2. ***(ii) Tier 2 (KL maximal lottery):*** *Minimizing $\max_{\pi'} \pi^\top M \pi' + \beta\, \mathrm{KL}(\pi \| m_x)$ on skew-symmetric $M(y, y') := P(y', y) - P(y, y')$ (Munos et al. 2024; Swamy et al. 2024) yields unique $\pi_\beta^{\text{ML}} \to \pi_0^\star = \arg\min_{\pi \in \mathcal{M}(P)} \mathrm{KL}(\pi \| m_x)$ as $\beta \to 0^+$, where $\mathcal{M}(P)$ is the maximal-lottery set (randomized unbeatable winners). By composition-consistency (Fishburn 1984; Brandl, Brandt, and Seedig 2016), whenever $\mathcal{M}(P_{\mathcal{C}})$ is unique (including any strict Condorcet winner), $\pi_0^\star$'s class shares resist m_x and clones; when $|\mathcal{M}(P_{\mathcal{C}})| > 1$ ($P \equiv 1/2 \implies \pi_0^\star = m_x$) or $\beta > 0$, raw anchoring still transmits clone inflation.* ***(iii) Tier 3 (Standing necessity and hierarchical rule):*** *Neither global homotopy ($\lim_{\beta\to\infty} \pi_\beta = m_x$) nor raw-anchored Gibbs policies with m_x-independent class rewards $U(C)$ resist clone splits shifting $m_x(C)$; finite-β between-class invariance therefore requires an m_x-independent standing prior $\rho \in \operatorname{int} \Delta(\mathcal{C})$ (extending Procaccia, Schiffer, and Zhang 2025). With strongly connected $G_+(\mathcal{C})$, quotient BT score $\hat{U}_\rho$ on $(\mathcal{C}, P_{\mathcal{C}}, \rho)$, within-class score u_C^{in}, and $C := C(y)$, the two-level policy ($\beta, \beta_{\text{in}} > 0$, normalizers Z_β, Z_C)*

$$\pi_\beta^{\text{hier}}(y) = \frac{\rho(C) e^{\hat{U}_\rho(C)/\beta}}{Z_\beta} \cdot \frac{m_x(y \mid C) e^{u_C^{\text{in}}(y)/\beta_{\text{in}}}}{Z_C} \qquad (4)$$

keeps between-class shares strictly invariant to m_x and clones for all $\beta > 0$, preserving within-class homotopy and preferences (Appendix C).

If loosening the leash pulls the whole policy to m_x, multiplying a stance's paraphrases pulls mass into its class. Eq. (4) replaces global homotopy with within-class homotopy ($m_x(\cdot \mid C)$) plus declared standing ρ. On Table 3 ($\mathcal{C} = \{\{a\}, \{b_1, b_2\}\}$, $\rho = (1/2, 1/2)$, $\beta = \beta_{\text{in}} = 1/2$), (4) restores a to $49/53 \approx 92.5\%$ (first-choice 65.3%) and lifts b_1 inside $\{b_1, b_2\}$ from 31.25% to 97.9%. Swapping $\hat{U}_\rho$ for ρ-anchored maximal lotteries guards $|\mathcal{C}| \geq 3$ against cross-class decoys, and Prop. C.1 (Appendix C) proves graceful stability under ε-near paraphrases ($\mathrm{TV} \leq \varepsilon / (4\beta s_{\min} \sqrt{\rho_{\min}})$).

5 Scope, Empirical Roadmap, and Implications

Five limits bound our claims: web corpora carry citation and synthetic dependence; deployed pipelines anchor to supervised fine-tuning (π_{SFT}); off-policy DPO weakens Vote 1 when pairs originate outside m_x; neural networks share cross-prompt features ϕ under non-convex training; and open-ended text yields near-clones, making $\mathcal{C}$ and ρ explicit institutional choices. Our results also cover closed single-turn slates; multi-step rollouts (tokenwise KL), interactive environments ($I(\mathcal{E}; S \mid \mathcal{D}, \mathcal{T}, \xi) > 0$), and long-horizon verifier search mark where static voting meets sequential control, without attributing subjective experience to silicon.

Three open-weight experiments probe beyond tabular theory. First, a *menu-vs.-leash ablation* crosses pair sampling ($\nu = m_x$ vs. ν_{bal}) against anchoring ($\pi_{\text{ref}} = m_x$ vs. uniform or class-normalized) on fixed ratings across β; cross-prompt spillover exceeding $|\Delta_{\text{unif}} - \Delta_{m_x}|/\beta$ falsifies tabular double voting as the main driver. Second, *decoy padding* injects weak variants near τ to test whether BT flips Condorcet winners while Tier 3 holds class shares. Third, a *pooling-order test* separates linear, log-odds, and raw-rating pooling ($\bar{P}$ for $|A| \geq 3$).

When a language model writes an unwritten consensus or overrides a four-in-five human majority, the reflex is to imagine an alien mind emerging inside silicon. Yet Table 3's reversal vanishes the moment the voting rule changes, with every human rating intact. Recognizing that pretraining votes twice turns alignment from mystique into social-choice architecture: balance the comparison menu, calibrate the leash, and declare who gets a vote in full daylight.

References

Abbas, A. E. 2009. A Kullback-Leibler View of Linear and Log-Linear Pools. *Decision Analysis*, 6(1): 25–37.

Banerjee, A.; Merugu, S.; Dhillon, I. S.; and Ghosh, J. 2005. Clustering with Bregman Divergences. *Journal of Machine Learning Research*, 6: 1705–1749.

Betz, G.; and Richardson, K. 2022. Judgment Aggregation, Discursive Dilemma and Reflective Equilibrium: Neural Language Models as Self-Improving Doxastic Agents. *Frontiers in Artificial Intelligence*, 5: 900943.

Bradley, R. A.; and Terry, M. E. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4): 324–345.

Brandl, F.; Brandt, F.; and Seedig, H. G. 2016. Consistent Probabilistic Social Choice. *Econometrica*, 84(5): 1839–1880.

Chakraborty, S.; Qiu, J.; Yuan, H.; Koppel, A.; Huang, F.; Manocha, D.; Bedi, A. S.; and Wang, M. 2024. MaxMin-RLHF: Alignment with Diverse Human Preferences. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235, 6116–6135.

Conitzer, V.; Freedman, R.; Heitzig, J.; Holliday, W. H.; Jacobs, B. M.; Lambert, N.; Mossé, M.; Pacuit, E.; Russell, S.; Schoelkopf, H.; Tewolde, E.; and Zwicker, W. S. 2024. Position: Social Choice Should Guide AI Alignment in Dealing with Diverse Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, 9346–9360.

Cover, T. M.; and Thomas, J. A. 2006. *Elements of Information Theory*. Wiley-Interscience, 2nd edition.

Dietrich, F.; and List, C. 2016. Probabilistic Opinion Pooling. In Hájek, A.; and Hitchcock, C., eds., *The Oxford Handbook of Probability and Philosophy*, 519–543. Oxford University Press.

Fishburn, P. C. 1984. Probabilistic Social Choice Based on Simple Voting Comparisons. *The Review of Economic Studies*, 51(4): 683–692.

Ford, L. R., Jr. 1957. Solution of a Ranking Problem from Binary Comparisons. *The American Mathematical Monthly*, 64(8P2): 28–33.

Galton, F. 1907. Vox Populi. *Nature*, 75(1949): 450–451.

Ge, L.; Halpern, D.; Micha, E.; Procaccia, A. D.; Shapira, I.; Vorobeychik, Y.; and Wu, J. 2024. Axioms for AI Alignment from Human Feedback. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37.

Grünwald, P. D. 2007. *The Minimum Description Length Principle*. MIT Press.

Hong, L.; and Page, S. E. 2004. Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers. *Proceedings of the National Academy of Sciences*, 101(46): 16385–16389.

Korbak, T.; Perez, E.; and Buckley, C. L. 2022. RL with KL Penalties is Better Viewed as Bayesian Inference. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, 1083–1091.

Krogh, A.; and Vedelsby, J. 1995. Neural Network Ensembles, Cross Validation, and Active Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 7, 231–238.

Munos, R.; Valko, M.; Calandriello, D.; Azar, M. G.; Rowland, M.; Guo, Z. D.; Tang, Y.; Geist, M.; Mesnard, T.; Michi, A.; Selvi, M.; Girgin, S.; Momchev, N.; Bachem, O.; Mankowitz, D. J.; Precup, D.; and Piot, B. 2024. Nash Learning from Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235, 36743–36768.

Procaccia, A. D.; Schiffer, B.; and Zhang, S. 2025. Clone-Robust AI Alignment. *arXiv preprint arXiv:2501.09254*.

Rafailov, R.; Sharma, A.; Mitchell, E.; Manning, C. D.; Ermon, S.; and Finn, C. 2023. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 53728–53741.

Redondo, N. 2026. The Alignment Problem Is an Identity Problem: Lessons from Four Billion Years of Major Evolutionary Transitions. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, volume 9, 2172–2181.

Siththaranjan, A.; Laidlaw, C.; and Hadfield-Menell, D. 2024. Distributional Preference Learning: Understanding and Accounting for Hidden Context in RLHF. In *The Twelfth International Conference on Learning Representations (ICLR)*.

Swamy, G.; Dann, C.; Kidambi, R.; Wu, Z. S.; and Agarwal, A. 2024. A Minimaximalist Approach to Reinforcement Learning from Human Feedback. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235, 47345–47377.

Tideman, T. N. 1987. Independence of Clones as a Criterion for Voting Rules. *Social Choice and Welfare*, 4(3): 185–206.

Vojnovic, M.; and Yun, S.-Y. 2025. What is the Alignment Objective of GRPO? *arXiv preprint arXiv:2502.18548*.

Xu, Y.; Zhao, S.; Song, J.; Stewart, R.; and Ermon, S. 2020. A Theory of Usable Information under Computational Constraints. In *The Eighth International Conference on Learning Representations (ICLR)*.

Zermelo, E. 1929. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29(1): 436–460.

A Proofs for Sections 2 and 3

Proof of Proposition 1. Allow optional $\mathcal{E}$ in $\Theta_T = F(\Theta_0, \mathcal{D}, \mathcal{T}, \mathcal{E}, \xi)$ with $I(\mathcal{E}; S \mid \mathcal{D}, \mathcal{T}, \xi) = 0$ ($S := S^{(N)}$). **(a)** Since $\xi \perp (\mathcal{D}, \mathcal{T}, S)$ and $\Theta_0 = g_0(\xi)$, data processing (DPI) gives $I(\Theta_0; S) \le I(\xi; S) = 0$ and $I(\Theta_0, \xi; S \mid \mathcal{D}, \mathcal{T}) \le I(\xi; S, \mathcal{D}, \mathcal{T}) = 0$. DPI and the chain rule give $I(\Theta_T; S) \le I(\mathcal{D}; S) + I(\mathcal{T}; S \mid \mathcal{D}) + I(\Theta_0, \xi; S \mid \mathcal{D}, \mathcal{T}) + I(\mathcal{E}; S \mid \mathcal{D}, \mathcal{T}, \Theta_0, \xi)$; the last two vanish, and Kraft gives $I(\mathcal{T}; S \mid \mathcal{D}) \le H(\mathcal{T}) \le \mathbb{E}[\ell(\mathcal{T})]$, yielding (1) and the $o(N)$ cap. **(b)** With $H_{\mathcal{V}}(S \mid Z) := \inf_{f \in \mathcal{V}} \mathbb{E}[-\log_2 f[Z](S)]$ and $I_{\mathcal{V}}(Z \to S) := H_{\mathcal{V}}(S \mid \varnothing) - H_{\mathcal{V}}(S \mid Z)$, optional ignorance gives $H_{\mathcal{V}}(S \mid \varnothing) = H(S)$ and Gibbs gives $H_{\mathcal{V}}(S \mid Z) \ge H(S \mid Z)$; DPI on $S - (X, \mathcal{D}, \mathcal{T}) - Z$ yields $I_{\mathcal{V}}(Z \to S) \le I(Z; S) \le I(X, \mathcal{D}, \mathcal{T}; S)$. For $\mathcal{V}_{\mathrm{lin}} = \{f_\theta(z)(1) = \sigma(\theta_0 + \theta^\top z)\}$ on the two-coin puzzle ($H(S) = I(X; S) = 1$), set $\delta_\pm := \theta_1 \pm \theta_2$ and convex $\ell(t) := \log_2(1 + e^t)$. On X, Jensen gives loss $(1/4)[\ell(-\theta_0 - \delta_+) + \ell(-\theta_0 + \delta_+) + \ell(\theta_0 + \delta_-) + \ell(\theta_0 - \delta_-)] \ge (1/2)[\ell(-\theta_0) + \ell(\theta_0)] \ge \ell(0) = 1$ (tight at $\theta = 0$), so $I_{\mathcal{V}_{\mathrm{lin}}}(X \to S) = 0$. On $Z = (Z_1, Z_2)$ ($Z_1 + Z_2 = 2S$), $\theta = (-c, c, c)$ has margin $(2S - 1)(-c + cZ_1 + cZ_2) = c$ and loss $\ell(-c) \to 0$ ($c \to \infty$), giving $I_{\mathcal{V}_{\mathrm{lin}}}(Z \to S) = 1$. For verifiers, $H(S) > 0$ and $H(S \mid \mathcal{D}, \mathcal{T}) = 0 \implies I(\mathcal{D}, \mathcal{T}; S) = H(S)$. □

Proof of Theorem 1. **(a)** For minimal regular $\mathcal{Q} = \{h(y) e^{\theta^\top \phi(y) - \psi(\theta)} : \theta \in \Omega\}$ ($\bar\mu_w \in \nabla\psi(\Omega)$) and finite $\mathrm{KL}(p_i \| q)$, $\sum_i w_i [\mathrm{KL}(p_i \| q) - \mathrm{KL}(p_i \| \bar p_w)] = \mathrm{KL}(\bar p_w \| q)$, so $\bar p_w$ minimizes cross-entropy. Minimability makes ψ strictly convex, fixing unique $\theta^\star$ ($q^\star := q_{\theta^\star}$) with $\nabla\psi(\theta^\star) = \bar\mu_w$. Since $\ln(q^\star / q_\theta)$ is affine in ϕ, its $\bar p_w$- and $q^\star$-means match, giving $\mathrm{KL}(\bar p_w \| q_\theta) = \mathrm{KL}(\bar p_w \| q^\star) + \mathrm{KL}(q^\star \| q_\theta)$ and (2). For $\mathcal{N}(2,1), \mathcal{N}(8,1), q^\star = \mathcal{N}(5,1), q = \mathcal{N}(6,1)$, doubling (2) yields $10 = 1.38 + 7.62 + 1$ via $2 \ln 2 - 2\mathbb{E}_{Z \sim \mathcal{N}(0,1)}[\ln(1 + e^{6Z - 18})] \approx 1.38$; Table 1 ($\phi_k = \mathbb{I}[y_k = Y]$, $q_{\mathrm{unif}} \equiv 1/8$) gives $\ln 8 = \ln 3 + \ln(9/4) + \ln(32/27)$. **(b)** $\pi^\star_\beta(y \mid x) > 0 = m_x(y)$ forces $\mathrm{KL} = \infty$; Bayes gives $\omega_i(x)$ for $\bar p_w(x) > 0$. □

Proposition A.1 (Mix-Share Contraction and Global Stability)**.** ***(a)*** *For $w, w' \in \operatorname{int} \Delta_K$ with $P(I{=}i) = w_i$, $P'(I{=}i) = w'_i$, $P(Y|i) = P'(Y|i) = p_i$, $\mathrm{KL}(w' \| w) - \mathrm{KL}(\bar p_{w'} \| \bar p_w) = \mathbb{E}_{\bar p_{w'}}[\mathrm{KL}(P'_{I|Y} \| P_{I|Y})] \ge 0$, equal iff $P'_{I|Y} = P_{I|Y}$ on $\operatorname{supp}(\bar p_{w'})$.* ***(b)*** *For w-independent P_X and minimal $q_\theta(y \mid x) = h(y \mid x) e^{\theta^\top \phi(x,y) - \psi_x(\theta)}$ with $\|\phi\|_2 \le R$ a.s., if $\theta^\star, \theta'$ are the conditional KL projections of $\bar p_w, \bar p_{w'}$ and $\mathbb{E}_{P_X}[\nabla^2 \psi_X(\theta)] \succeq \lambda I$ ($\lambda > 0$) on $[\theta^\star, \theta']$, then $\mathbb{E}_{P_X}[\mathrm{TV}(q_{\theta^\star}, q_{\theta'})] \le (R/(2\sqrt{\lambda}))\|w' - w\|_1$.*

Proof. Expanding $\mathrm{KL}(P'_{I,Y} \| P_{I,Y})$ both ways gives (a). For (b), optimality yields $\mathbb{E}_{P_X}[\nabla \psi_X(\theta^\star)] = \bar\mu_w := \mathbb{E}_{P_X, \bar p_w}[\phi]$ (likewise $\theta', \bar\mu_{w'}$), so $D_{\mathrm{sym}} := \mathbb{E}_{P_X}[\mathrm{KL}(q_{\theta^\star} \| q_{\theta'}) + \mathrm{KL}(q_{\theta'} \| q_{\theta^\star})] = (\theta^\star - \theta')^\top (\bar\mu_w - \bar\mu_{w'}) \ge \lambda \|\theta^\star - \theta'\|_2^2$. Since $\|\bar\mu_w - \bar\mu_{w'}\|_2 \le R \|w' - w\|_1$, Cauchy-Schwarz gives $D_{\mathrm{sym}} \le (R^2/\lambda)\|w' - w\|_1^2$; Jensen and Pinsker ($4\,\mathrm{TV}^2 \le D_{\mathrm{sym}}$) conclude. □

B Proofs for Sections 4.1 and 4.2

Proof of Theorem 2. **(a)** For symmetric $N_{yy'} \ge 0$ and $\delta_{yy'} := u(y) - u(y')$, pairing $y < y'$ gives $\partial_{u(y)} \mathcal{L}_{\mathrm{BT}}(u) = \sum_{y' \ne y} N_{yy'} [P(y, y') - \sigma(\delta_{yy'})]$ and $-v^\top \nabla^2 \mathcal{L}_{\mathrm{BT}}(u) v = \sum_{y < y'} N_{yy'} \sigma'(\delta_{yy'}) (v(y) - v(y'))^2$. Strong connectivity of G_+ implies coercivity on $U_0 = \{u : \sum_y u(y) = 0\}$ and unique $\hat u_N \in U_0$. (i) If $P(y, y') = \sigma(u^\star(y) - u^\star(y'))$, $\nabla \mathcal{L}_{\mathrm{BT}}(u^\star) = 0 \implies \hat u_N = u^\star$. (ii) For $N_{yy'} = \nu(y)\nu(y')$, dividing $\partial_{u(y)} \mathcal{L}_{\mathrm{BT}}(\hat u_\nu) = 0$ by $\nu(y) > 0$ and adding $\nu(y)/2$ yields $\sum_{y'} \nu(y') \sigma(\hat u_\nu(y) - \hat u_\nu(y')) = B_\nu(y)$, giving strict monotonicity; $0 < \sigma' \le 1/4$ and MVT give $|B_\nu(y_1) - B_\nu(y_2)| \le |\hat u_\nu(y_1) - \hat u_\nu(y_2)|/4$. (iii) On Table 2 ($N_{ac} = N_{bd} = 20$, others 1; wins $12.40, 17.90, 9.80, 3.90$), write $\sigma_{yz} := \sigma(\hat u_N(y) - \hat u_N(z))$. If $r_a \ge r_b$, b's wins (17.90) with $\sigma_{ba} \le .50, \sigma_{bc} < 1$ force $\sigma_{bd} > .82$, so $r_a / r_d \ge r_b / r_d > 41/9$ ($\sigma_{ad} > .82$). Then a's wins (12.40) with $\sigma_{ab} \ge .50$ force $\sigma_{ac} < .554$ ($\sigma_{ca} > .446$), $r_c / r_b \ge r_c / r_a > 223/277$ ($\sigma_{cb} > .446$), and $r_c / r_d > (223/277)(41/9) = 9143/2493$ ($\sigma_{cd} > .785$), making c's wins $> 20(.446) + .446 + .785 = 10.151 > 9.80$, contradiction; hence $r_b > r_a$. Also $r_b \le r_d$ makes b's wins $< 12 < 17.90$, and $r_c \ge r_b > r_a$ makes c's wins $> 10.50 > 9.80$; thus $\hat u_N(b)$ is uniquely largest. **(b)** Set $B_y := B_{m_x}(y)$, $m_0 := m_x(a)$, $m_k := m_x(b_k)$. Since $\alpha_1, \alpha_2 \in (1/2, 1)$ and $m_x \in \operatorname{int} \Delta(A)$, G_+ stays strongly connected even at $\gamma = 1$ via $b_2 \to a \to b_1$ (transitive iff $\alpha_1 - \alpha_2 \le 1 - \gamma$). Algebra gives $B_a - B_{b_1} = (\alpha_1 - 1/2) - [(\alpha_1 - 1/2) + (\gamma - \alpha_2)] m_2$ and $B_a - B_{b_2} = (\alpha_2 - 1/2)(1 - m_1) + (\alpha_1 + \gamma - 1) m_1 > 0$. Thus $\gamma \le \alpha_2$ or $m_2 < \tau$ gives $\hat u_{m_x}(a) > \max\{\hat u_{m_x}(b_1), \hat u_{m_x}(b_2)\}$ ($\pi^\star_\beta(a) \to 1$), while $\gamma > \alpha_2$ and $m_2 > \tau$ gives $\pi^\star_\beta(b_1) \to 1$. At $m_2 = \tau$, $B_a = B_{b_1} > B_{b_2}$ gives $\hat u_{m_x}(a) = \hat u_{m_x}(b_1)$ and $\pi^\star_\beta(b_1) \to m_1/(m_0 + m_1) < 1$ (strict iff). With $\bar p_{-j}(x) := \sum_{i \ne j} w_i p_i(x) / (1 - w_j)$, $m_2 = \omega_j(x) e_j = e_j w_j \Lambda_j(x) / [w_j \Lambda_j(x) + 1 - w_j] > \tau \iff w_j > \tau / [\tau + (e_j - \tau)\Lambda_j(x)]$. □

Verification of Table 3 and Proof of Theorem 3.
Table 3: Uniform ν with $r = (8, 4, 1)$ matches a's wins ($8/12 + 8/9 = 14/9$) and b_2's ($1/9 + 1/5 = 14/45$); $\nu = m_x = (13, 40, 88)/141$ with $r = (8, 7, 1)$ matches 141-scaled wins for a ($40(8/15) + 88(8/9) = 896/9 = 128(7/9)$) and b_2 ($13/9 + 40/8 = 58/9 = 13(2/9) + 40(4/45)$). At $\beta = 1/2$, log-odds shift from $\ln 4 = 1.386$ by $\ln(16/49) = -1.119$ (Vote 1) and $\ln(13/40) = -1.124$ (Vote 2) to $-0.857 = \ln(104/245)$; $m_x(b_2) = 88/141$ at $e_j{=}1, \Lambda_j{=}100$ uses $w_j = 22/1347 \approx 1.63\%$. **Thm. 3(a):** Let $m_0 := m_x(a)$, $m_k := m_x(b_k)$, $\Delta^{yz}_\nu := \hat u_\nu(y) - \hat u_\nu(z)$, $\ln_+(t) := \max(0, \ln t)$ ($c/0 := \infty$). Since $B_\nu(b_1) - B_\nu(b_2) = (\gamma - 1/2)(\nu(b_1) + \nu(b_2)) > 0$, $\hat u_\nu(b_1) > \hat u_\nu(b_2)$ for all full-support ν, and $m_2 < \tau$ keeps a unique mode in Row 2 for all $\beta > 0$. Set $\beta_{\mathrm{mode}} := \min\{\Delta^{a,b_2}_{\mathrm{unif}} / \ln_+(m_2/m_0), \Delta^{b_1,b_2}_{m_x} / \ln_+(m_2/m_1)\}$. Row 3 keeps a unique mode iff $\beta < \min(\Delta_{\mathrm{unif}}/\kappa, \Delta^{a,b_2}_{\mathrm{unif}} / \ln_+(m_2/m_0))$; Row 4 makes b_1 unique mode iff $\Delta_{m_x}/\kappa < \beta < \Delta^{b_1,b_2}_{m_x} / \ln_+(m_2/m_1)$. Inter-

secting gives $(\Delta_{m_x}/\kappa, \min(\Delta_{\mathrm{unif}}/\kappa, \beta_{\mathrm{mode}}))$, nonempty iff $\Delta_{m_x}/\kappa < \beta_{\mathrm{mode}}$ (guaranteed as $m_2 \uparrow \tau$ at fixed $\kappa > 0$, since $\Delta_{m_x} \to 0^+$ while $\Delta^{b_1,b_2}_{m_x} \geq 4(\gamma - 1/2)\tau > 0$). Table 3 yields $(\ln(8/7)/\ln(40/13), \ln 2/\ln(40/13)) \approx (0.119, 0.617)$, $\beta_{\mathrm{mode}} = \min\{\ln 8/\ln(88/13), \ln 7/\ln(88/40)\} \approx 1.087$.

Thm. 3(b): Set $K := 2/\eta > 8$ and $m^{(t)} = ((1-t)/(1+K^3), K^3(1-t)/(1+K^3), t)$ on $t \in (0, \tau]$ ($\kappa = 3\ln K$). Since $\tau > 1/3$, $B_{\mathrm{unif}}(a) - B_{\mathrm{unif}}(b_1) = (\gamma - 1/2)(\tau - 1/3) > 0$, so $\Delta^{a,b_2}_{\mathrm{unif}} > \Delta_{\mathrm{unif}} > 0$. On $(0, \tau]$, $B_{m^{(t)}}(b_1) - B_{m^{(t)}}(b_2) = (\gamma - 1/2)(1 - (1-t)/(1+K^3)) > (\gamma - 1/2)/2$, so Thm. 2(a)(ii) gives $\Delta^{b_1,b_2}_{m^{(t)}} > 2(\gamma - 1/2)$. Let $M := \min\{\Delta_{\mathrm{unif}}, 2(\gamma - 1/2)\} > 0$, $q_\tau := 2\tau/(1-\tau)$, and $\beta := M/[4\ln K + \ln_+ q_\tau]$. Using $(1+K^3)/K^3 < 2$, for all $t \in (0, \tau)$: Row 3 has $\pi^{(3)}(b_1)/\pi^{(3)}(a) \leq K^3 e^{-4\ln K} = K^{-1}$ and $\pi^{(3)}(b_2)/\pi^{(3)}(a) < q_\tau K^3 e^{-\Delta_{\mathrm{unif}}/\beta} \leq K^{-1}$ ($\pi^{(3)}(a) > 1/(1+\eta) > 1 - \eta$); Row 4 has $\pi^{(4)}(b_2)/\pi^{(4)}(b_1) < q_\tau e^{-2(\gamma-1/2)/\beta} \leq K^{-4} < K^{-1}$; Row 2 has $\pi^{(2)}(b_2)/\pi^{(2)}(a) < e^{-4\ln K} < K^{-1}$. Since $-\nabla^2 \mathcal{L}_{\mathrm{BT}} \succ 0$ on U_0, IFT makes $\Delta(t) := \Delta_{m^{(t)}}$ continuous on $[\tau/2, \tau]$ with $\Delta(\tau) = 0$ and $\Delta(\tau/2) \geq 2\tau(\gamma - 1/2) > (2/3)(\gamma - 1/2) > \beta \ln K$. By IVT, some $t^\star \in (\tau/2, \tau)$ obeys $\Delta(t^\star) = \beta \ln K$. At $m_x = m^{(t^\star)}$, Row 2 has $\pi^{(2)}(b_1)/\pi^{(2)}(a) = K^{-1}$ ($\pi^{(2)}(a) > 1 - \eta$) and Row 4 has $\pi^{(4)}(a)/\pi^{(4)}(b_1) = K^{-3} e^{\ln K} = K^{-2} < K^{-1}$, so $\pi^{(4)}(b_1) > 1 - \eta$ and $\pi^{(4)}(a) < \eta$. □

C GRPO, Pooling Order, and Tier 3 Proofs

GRPO and pooling order. For fixed standardized advantage $\tilde{R}$, the stop-gradient Schulman reverse-KL objective $\mathcal{J}(\pi; \pi_{\mathrm{old}}) = \sum_y \pi(y)\tilde{R}(y) - \beta \sum_y \pi_{\mathrm{old}}(y)[m_x(y)/\pi(y) - \ln(m_x(y)/\pi(y)) - 1]$ satisfies $\partial_{\pi(y)}\mathcal{J} = \lambda$ at $\pi_{\mathrm{old}} = \pi$ iff $\tilde{R}(y) + \beta(m_x(y)/\pi(y) - 1) = \lambda$, giving $\pi^\star_{\mathrm{GRPO}}(y) = \beta m_x(y)/(\lambda + \beta - \tilde{R}(y))$. For $m_x \in \operatorname{int}\Delta(A)$, $\sum_y \pi^\star_{\mathrm{GRPO}}(y)$ decreases strictly from ∞ to 0 on $(\max_y \tilde{R}(y) - \beta, \infty)$, fixing unique $\lambda^\star = \mathbb{E}_{\pi^\star_{\mathrm{GRPO}}}[\tilde{R}]$. For three equal groups at $(24/25, 1/4, 1/4)$ (odds $24, 1/3, 1/3$) on $\{a, b\}$: linear pooling gives $73/150 \approx 48.7\%$ (b wins), log-odds pooling gives $\bar{\delta} = (1/3)\ln(8/3) > 0$ and odds $(8/3)^{1/(3\beta)} > 1$ (58.1% at $\beta = 1$, a wins), and raw rating pooling gives $\bar{P}(a, b) = 73/150$ ($|A| \geq 3$ separates all three).

Proof of Theorem 4. **(i)** Immediate from m_x-independence of ν_{bal} and strong connectivity of $G_+(\nu_{\mathrm{bal}})$. **(ii)** With $M^\top = -M$, $f_\beta(\pi) := \max_{\pi' \in \Delta(A)} \pi^\top M \pi' + \beta\,\mathrm{KL}(\pi \| m_x)$ is continuous and strictly convex on compact $\Delta(A)$, fixing unique π^{ML}_β. Skew-symmetry makes $\mathcal{M}(P) = \{\pi : \max_{\pi'} \pi^\top M \pi' \leq 0\}$ nonempty, closed, convex, with unique KL-minimizer $\pi^\star_0$, and gives $0 \leq \max_{\pi'} (\pi^{\mathrm{ML}}_\beta)^\top M \pi' \leq \beta\,\mathrm{KL}(\pi^\star_0 \| m_x) \to 0$ and $\mathrm{KL}(\pi^{\mathrm{ML}}_\beta \| m_x) \leq \mathrm{KL}(\pi^\star_0 \| m_x)$, forcing $\pi^{\mathrm{ML}}_\beta \to \pi^\star_0$ by compactness and uniqueness. For $\pi \in \mathcal{M}(P)$ with class marginal $\Pi(C) := \sum_{y \in C} \pi(y)$, lifting $Q \in \Delta(\mathcal{C})$ via $\pi_Q(y) := Q(C)\pi(y)/\Pi(C)$ if $\Pi(C) > 0$ (arbitrary else) cancels within-class sums by skew-symmetry, giving $0 \geq \pi^\top M \pi_Q = \Pi^\top M_{\mathcal{C}} Q$ and $\Pi \in \mathcal{M}(P_{\mathcal{C}})$; thus unique $\mathcal{M}(P_{\mathcal{C}})$ makes $\pi^\star_0$'s class marginals invariant to m_x and clones. Conversely: (1) $P \equiv 1/2 \implies \pi^\star_0 = m_x$; (2) on $\{a\}, \{b_1, b_2\}$ with $P_{\mathcal{C}} \equiv 1/2$ and $P(b_1, b_2) > 1/2$, $\pi^\star_0(a) = m_x(a)/(m_x(a) + m_x(b_1))$; (3) at $\beta > 0$ with strict Condorcet a ($P(a, b_j) = \alpha > 1/2$, $P(b_j, b_\ell) = 1/2$ in C_2), $\max_{\pi'} \pi^\top M \pi' = (1 - \pi(a))(2\alpha - 1)$, so the KL chain rule gives $\pi^{\mathrm{ML}}_\beta(a)/(1 - \pi^{\mathrm{ML}}_\beta(a)) = (m_x(C_1)/m_x(C_2))e^{(2\alpha-1)/\beta}$, varying strictly with $m_x(C_2)$. **(iii)** Global homotopy $\pi_\beta \to m_x$ forces $\Pi_\beta(C) \to m_x(C)$, and raw-anchored Gibbs $\Pi_\beta(C) \propto m_x(C)e^{U(C)/\beta}$ varies for all $\beta > 0$. For (4), let $\pi^{\mathrm{hier}}_{\mathrm{ref}}(y) := \rho(C(y))m_x(y \mid C(y))$ and maximize $\mathbb{E}_\pi[\hat{U}_\rho(C(y)) + (\beta/\beta_{\mathrm{in}})u^{\mathrm{in}}_{C(y)}(y) - \beta \ln Z_{C(y)}] - \beta\,\mathrm{KL}(\pi \| \pi^{\mathrm{hier}}_{\mathrm{ref}})$. By the KL chain rule, inner maximization gives $\pi^{\mathrm{in}}_{\beta_{\mathrm{in}}}(\cdot \mid C)$ with value $+\ln Z_C$, canceling $-\ln Z_C$ and leaving $\Pi^{\mathrm{out}}_\beta(C) = \rho(C)e^{\hat{U}_\rho(C)/\beta}/Z_\beta$, independent of m_x and clones. On Table 3 ($\rho = (1/2, 1/2), \beta = \beta_{\mathrm{in}} = 1/2$): $P_{\mathcal{C}}(C_1, C_2) = 7/9$ gives odds $(7/2)^2 = 49/4$ ($\Pi^{\mathrm{out}}_\beta(C_1) = 49/53 \approx 92.5\%$); inside C_2, $P(b_1, b_2) = 41/45$ and $m_x(b_1 \mid C_2) = 5/16$ give odds $(5/11)(41/4)^2 = 8405/176$ ($8405/8581 \approx 97.9\%$, first-choice 65.3%). □

Proposition C.1 (ε-Approximate Clone Stability for Tier 3)**.** *If $|P(y, y') - P_{\mathcal{C}}(C, C')| \leq \varepsilon$ for all $C \neq C', y \in C, y' \in C'$, set $\tilde{P}_{\mathcal{C}}(C, C') := \sum_{y \in C, y' \in C'} q(y \mid C)q(y' \mid C')P(y, y')$ for any $q(\cdot \mid C) \in \Delta(C)$. If $G_+(\mathcal{C}, P_{\mathcal{C}}), G_+(\mathcal{C}, \tilde{P}_{\mathcal{C}})$ are strongly connected with BT scores $\hat{U}_\rho, \tilde{U}_\rho$, $s_{\min} := \min_{t \in [0,1],\, C \neq C'} \sigma'((1-t)(\hat{U}_\rho(C) - \hat{U}_\rho(C')) + t(\tilde{U}_\rho(C) - \tilde{U}_\rho(C'))) > 0$, and $\rho_{\min} := \min_C \rho(C) > 0$, then $\mathrm{TV}(\tilde{\Pi}^{\mathrm{out}}_\beta, \Pi^{\mathrm{out}}_\beta) \leq \varepsilon/(4\beta\, s_{\min}\sqrt{\rho_{\min}})$.*

Proof. Let $\delta := \tilde{U}_\rho - \hat{U}_\rho$ and $S_2 := \sum_{C < C'} \rho(C)\rho(C')(\delta(C) - \delta(C'))^2 = \mathrm{Var}_\rho(\delta)$; $|\tilde{P}_{\mathcal{C}} - P_{\mathcal{C}}| \leq \varepsilon$. From $\nabla\mathcal{L}_{\mathrm{BT}}(\hat{U}_\rho; P_{\mathcal{C}}) = \nabla\mathcal{L}_{\mathrm{BT}}(\tilde{U}_\rho; \tilde{P}_{\mathcal{C}}) = 0$, pairing $C < C'$ with MVT on $[\hat{U}_\rho, \tilde{U}_\rho]$ and Cauchy-Schwarz gives $s_{\min} S_2 \leq \varepsilon\sqrt{(1 - \sum_C \rho(C)^2)/2}\sqrt{S_2} < (\varepsilon/\sqrt{2})\sqrt{S_2}$, so $S_2 \leq \varepsilon^2/(2s^2_{\min})$. With $a := \max\delta - \bar{\delta}_\rho \geq 0$, $b := \bar{\delta}_\rho - \min\delta \geq 0$, and $\mathrm{osc}(\delta) := a + b$, $S_2 \geq \rho_{\min}(a^2 + b^2) \geq (\rho_{\min}/2)\mathrm{osc}(\delta)^2$, so $\mathrm{osc}(\delta) \leq \varepsilon/(s_{\min}\sqrt{\rho_{\min}})$. Along $\Pi_t(C) \propto \rho(C)e^{(\hat{U}_\rho(C) + t\delta(C))/\beta}$, Popoviciu gives $D_{\mathrm{sym}}(\tilde{\Pi}^{\mathrm{out}}_\beta, \Pi^{\mathrm{out}}_\beta) = \beta^{-2}\int_0^1 \mathrm{Var}_{\Pi_t}(\delta)\,dt \leq \mathrm{osc}(\delta)^2/(4\beta^2)$; Pinsker ($4\,\mathrm{TV}^2 \leq D_{\mathrm{sym}}$) concludes. □