WHEN ARE ITEMWISE REWARDS ENOUGH? FINITE-SAMPLE AUDITS FOR RELATIONAL PREFERENCES

Anonymous authors
Paper under double-blind review

ABSTRACT

Rubrics specify what to evaluate; representing pairwise evidence within each criterion remains a separate choice. A prompt–rubric comparison table may remain relational or be summarized by one score per response, imposing Bradley–Terry differences. We make this choice testable at prespecified approximation tolerances. A simultaneous finite-sample audit labels each row scalar-adequate, nonintegrable, or indeterminate, allowing itemwise and relational representations to coexist. For a fixed full-flow estimator, Hodge orthogonality, row-additive squared loss, and Cartesian actions reduce an explicit 2^R-action source-risk problem to R thresholds: projection pays squared cycle bias and retention pays cyclic estimation MSE. Independent audit and recovery blocks yield regret controlled by threshold-straddling intervals. Component-specific low-rank recovery and entropy-regularized Nash stability then produce policy certificates. In paired Bernoulli experiments, the selector lowers weighted source risk by 38.9% versus forced scalar and 23.4% versus forced relational representation, with no observed unsafe projection relative to the independently calibrated registry oracle. Separately, a continuous odd edgewise transform preserves every score-difference flow exactly when linear; expected-win margins can therefore make scalar source logits nonadditive. In 3,219 human votes across 183 exact Summarize-from-Feedback cells, all prompts remain indeterminate over the examined tolerances. In 3,115 UltraFeedback records, nonlinear scalar-score aggregation produces nonadditive payoffs with measurable policy effects. The resulting method projects only certified rows and retains relational structure when the audit is indeterminate.

1 INTRODUCTION

Rubrics specify what an evaluator should judge; representation within each criterion is a separate design choice. For a fixed prompt and rubric, a system may retain an antisymmetric table of pairwise log-odds or summarize it by one score per response. Current alignment pipelines use both designs: rubric scores can be aggregated into rewards, while criterion-wise pairwise judges retain relational comparisons (Gunjal et al., 2026; Jia et al., 2026). Itemwise scores provide a compact interface for ranking and policy optimization (Christiano et al., 2017; Ouyang et al., 2022; Rafailov et al., 2023). For m responses, they reduce $\binom{m}{2}$ edge values to $m-1$ degrees of freedom. We ask when finite evidence warrants this simplification at the resolution relevant to the downstream decision.

For prompt q, rubric r, and responses i, j, let $\theta_{qr,ij}$ be the population preference log-odds. Bradley–Terry representability requires $\theta_{qr,ij} = u_{qr,i} - u_{qr,j}$ (Bradley & Terry, 1952). The rubric index remains intact throughout; the audited simplification replaces a relational table within one criterion by an itemwise potential. A strict three-way cycle violates this representation through its edge signs. A magnitude inconsistency can be equally decisive without a sign cycle: $\theta_{AB} = \theta_{BC} = 1$ and $\theta_{AC} = 1.4$ preserve $A > B > C$ while violating additivity. Finite samples can create apparent circulation, and a population residual can be too small to matter for a declared policy tolerance.

HodgeRank separates gradient and cycle flows (Jiang et al., 2011); low-rank skew models recover nontransitive comparisons (Yang & Wakin, 2015; Chen & Joachims, 2016; Lee & Chen, 2026); and composite tests study Bradley–Terry and generalized Thurstone fit (Makur & Singh, 2025a;b). Nash methods optimize general preference relations (Munos et al., 2024; Swamy et al., 2024). Attention-scaled Hodge diagnostics with a parametric-bootstrap scalarity test (Xing, 2026) and

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

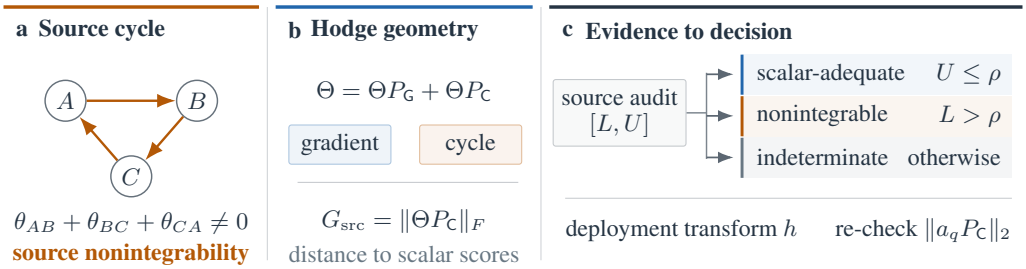


Figure 1: Nonzero source-logit circulation witnesses failure of exact scalar representation; Hodge projection also detects magnitude inconsistencies without a sign cycle. Confidence bounds $[L, U]$ compare the source residual with tolerance ρ to select one of three audit outcomes. Linear logit deployment preserves source geometry; a chosen nonlinear map, including expected-win semantics, requires a separate check of $\|a_q P_{\mathsf{C}}\|_2$.

decomposition-based dynamic Nash optimization (Huang et al., 2026) address adjacent links. These strands leave a selective decision open: exact-equality tests target zero tolerance, while recovery guarantees require stability analysis to translate estimation error into policy perturbation. Our three-way audit certifies projection eligibility, required relational retention, or unresolved evidence; within the eligible branch, a source-risk gate selects projection or retention. An unresolved row defaults to relational retention, while a prespecified operational rule may acquire comparisons or abstain.

Our adaptive representation audit stacks prompt–rubric edge-logit rows in Θ and projects them onto gradient and cycle spaces with P_{G} and P_{C}. The representation decision is whether to estimate each cyclic component or project it away. The residual $g_{qr} = \|\theta_{qr} P_{\mathsf{C}}\|_2$ is both distance to an itemwise potential and the exact scalar-bias term, while retaining the full row pays cyclic estimation MSE. For R rubrics and one fixed full-flow estimator, the distribution-dependent oracle minimizes weighted source-logit risk over all 2^R scalar/relational actions. Hodge orthogonality removes the within-row cross term; row-additive squared loss and the Cartesian action class remove cross-rubric coupling. Given the oracle inputs, the finite program becomes R closed-form comparisons, while policy-calibrated source-gap margins restrict feasible projections. Split audit and recovery samples estimate the constrained action with excess risk confined to threshold-straddling intervals. A two-component Hodge low-rank estimator recovers the full source table under full-coverage or design-conditional sparse assumptions; the risk threshold and source-gap margin answer the distinct questions of statistical efficiency and operational adequacy.

Deployment semantics introduce a separate choice. Linear source-logit payoffs preserve gradient geometry. Pairwise-win objectives use $\tanh(t/2) = 2\sigma(t) - 1$, the expected margin of a $\{+1, -1\}$ comparison outcome. This bounded, nonlinear map can change the geometry, and our characterization identifies linear maps as the continuous odd transforms that preserve every score-difference flow. We audit the chosen payoff geometry independently of source representability. Every selected row, including the relational fallback for an indeterminate row, passes through the declared deployment map and entropy-regularized Nash rule. On additive payoffs the rule is a reference-weighted softmax that concentrates on top-score responses as $\tau \downarrow 0$; on relational payoffs it computes the regularized equilibrium of the full game. Recovery and stability bounds provide global and local policy certificates; a prespecified operational rule may replace the fallback with acquisition or abstention. Figure 1 summarizes the two representation choices and the decision.

Contributions. **First**, we formulate within-rubric representation as a finite adaptive decision, solve its 2^R-action source-risk program in closed form, prove an audit-calibrated constrained-oracle inequality, and control simultaneous three-way errors (Sections 3 and 4). **Second**, we characterize the deployment boundary: continuous edgewise transforms preserve all score differences exactly when they are linear, and the common Nash interface recovers scalar ranking on additive payoffs (Section 3). **Third**, component-specific Hodge low-rank recovery and global, local, and hybrid policy certificates cover explicit full-coverage and design-conditional sparse regimes (Section 4). **Fourth**, theory-matched simulations and two data analyses test distinct links: exact-cell human evidence remains indeterminate at the examined policy tolerances, while UltraFeedback quantifies geometry and policy changes under chosen nonlinear payoff semantics (Section 5).

2 Related Work and Statistical Distinction

Preference representation and adaptive representation selection. Classical Bradley–Terry–Luce models assign one log-score to each item (Bradley & Terry, 1952; Luce, 1959). Recent matrix and tensor completion methods share low-rank information across tasks while retaining this scalar BTL premise (Li et al., 2026a;b); they occupy the $\Theta P_{\mathsf{C}} = 0$ scalar-BTL branch. Preference-representation limits and generalized Thurstone tests clarify reward representability and goodness of fit (Liu et al., 2026; Makur & Singh, 2025a;b). Our target is the selective decision at positive, policy-calibrated tolerance: certify which prompt–rubric rows support the scalar premise, retain the remaining relations, and propagate unresolved uncertainty. HodgeRank supplies the orthogonal gradient–cycle geometry (Jiang et al., 2011); low-rank skew models compress nontransitive relations (Yang & Wakin, 2015; Chen & Joachims, 2016; Lee & Chen, 2026); and Bayesian Hodge models incorporate covariates (Okahara et al., 2026). We add separate component penalties, a finite-sample indeterminate branch, and a source-risk oracle that selects rowwise projection before downstream decision making.

Matrix completion estimates a fixed representation. Classical matrix completion fixes a low-rank real latent matrix and studies recovery from missing or noisy entries (Candès & Recht, 2009; Koltchinskii et al., 2011); one-bit completion observes linked binary measurements of that matrix (Davenport et al., 2014). Nuclear-norm theory supplies decomposability, general-sampling tools, and restricted-curvature conditions (Negahban & Wainwright, 2011; 2012; Klopp, 2014). Our recovery lane applies these tools separately to ΘP_{G} and ΘP_{C}, with distinct ranks and penalties. The audit then certifies rowwise projection eligibility, the risk gate selects projection or retention, and the error propagates to payoff, policy, and exploitability. A single penalty on Θ gives unstructured logistic one-bit completion; imposing $\Theta P_{\mathsf{C}} = 0$ yields its gradient-constrained low-rank-BTL branch. Sparse recovery uses a separate RSC-with-tolerance condition on the Hodge error cone.

LLM alignment, games, and rubrics. General preference models and explicit transitive–cyclic decompositions extend LLM feedback beyond scalar rewards (Zhang et al., 2025; Huang et al., 2026). Nash learning and self-play provide decisions for general preference relations (Munos et al., 2024; Swamy et al., 2024; Wu et al., 2025); related theory studies exploration-dependent regret, Condorcet and Smith consistency, and maximal lotteries (Nan et al., 2026; Shi et al., 2025; Maura-Rivero et al., 2025). Weighted Hodge analysis with a parametric-bootstrap scalarity test on aggregated Arena votes (Xing, 2026) tests model-pair scalarity at the aggregate level. Rubric judges expose criterion-level evidence (Hashemi et al., 2024; Kim et al., 2024; Roy et al., 2026; Yang et al., 2026). Their outputs enter training through scalar reward aggregation (Gunjal et al., 2026) or criterion-wise pairwise comparison with external aggregation (Jia et al., 2026). Rubric-based systems can preserve criteria separately. Our question is finer: within each fixed criterion, does finite pairwise evidence support an itemwise score at the resolution required by the decision? We connect that audit to component-specific recovery and global, local, and hybrid policy certificates. Pooled observations identify aggregate geometry; individual attribution requires rater-stratified designs.

3 Rubric-Conditioned Edge Geometry

For prompt $q \in [Q]$, rubric $r \in [R]$, and $m \geq 3$ candidate responses, orient the complete edge set $\mathcal{E} = \{(i,j) : i < j\}$, $E = \binom{m}{2}$. A comparison $Y_{qre} \in \{0,1\}$ satisfies $\mathbb{P}(Y_{qre} = 1) = \sigma(\theta_{qre})$, where $\sigma(t) = (1 + e^{-t})^{-1}$. Set $\theta_{qr,ji} = -\theta_{qr,ij}$ and $\theta_{qr,ii} = 0$. Stacking $x = (q,r)$ gives $\Theta \in \mathbb{R}^{Z \times E}$ with $Z = QR$. Let $B \in \mathbb{R}^{E \times m}$ be the oriented incidence matrix with $B_{(i,j),k} = \mathbf{1}\{k = i\} - \mathbf{1}\{k = j\}$, $\mathsf{G} = \mathrm{range}(B)$ the gradient space, and $\mathsf{C} = \ker(B^\top)$ the cycle space. Their orthogonal projectors are $P_{\mathsf{G}} = BB^\dagger$ and $P_{\mathsf{C}} = I - P_{\mathsf{G}}$. Every row has the unique split

$$\Theta = \Theta_{\mathsf{G}} + \Theta_{\mathsf{C}}, \qquad \Theta_{\mathsf{G}} = \Theta P_{\mathsf{G}}, \quad \Theta_{\mathsf{C}} = \Theta P_{\mathsf{C}}. \tag{3.1}$$

The composite scalar null is $H_0 : \Theta P_{\mathsf{C}} = 0$, a stronger requirement than edge-sign transitivity. On a complete graph it is equivalent to zero circulation on every triangle; on an incomplete graph the full cycle space includes harmonic flow around holes.

The rubric and representation axes remain distinct. For $x = (q,r)$, define the within-rubric representation gap $g_x(\Theta) = \|\theta_x P_{\mathsf{C}}\|_2$ and its omnibus counterpart $G_{\mathrm{src}}(\Theta) = \|\Theta P_{\mathsf{C}}\|_F = \{\sum_x g_x(\Theta)^2\}^{1/2}$. The rowwise gap asks whether one fixed criterion admits an itemwise potential; the global gap

provides an omnibus summary. Rowwise decisions can therefore preserve the full rubric vector while mixing itemwise and relational representations across its coordinates.

We allow the two components to share information differently across prompt–rubric rows, with $\operatorname{rank}(\Theta_{\mathsf{G}}) \le r_{\mathsf{G}}$ and $\operatorname{rank}(\Theta_{\mathsf{C}}) \le r_{\mathsf{C}}$. Deployment utility is a modeling choice. Fix a declared globally L_h-Lipschitz odd map $h : \mathbb{R} \to \mathbb{R}$ with $L_h > 0$ and weights $w_{qr} \ge 0$, $\sum_r w_{qr} = 1$:

$$A_{q,ij} = \sum_{r=1}^{R} w_{qr} h(\theta_{qr,ij}), \qquad A_q = -A_q^{\top}. \tag{3.2}$$

Identity logit utility has $h(t) = t$ and $L_h = 1$. The bounded expected-win interface encodes a win as $+1$ and a loss as -1, giving $h(t) = 2\sigma(t) - 1 = \tanh(t/2)$ and $L_h = 1/2$; it is our primary nonlinear deployment rule. For the oriented edge vector $a_q = (A_{q,e})_{e\in\mathcal{E}}$, define the deployment residual $G_{\mathrm{dep}}(q;\Theta) = \|a_q P_{\mathsf{C}}\|_2$, separately from the source gaps above.

Proposition 3.1 (Universal scalarity-preserving transforms). *For any $h : \mathbb{R} \to \mathbb{R}$, the following are equivalent: (i) entrywise application of h maps every scalar flow on K_3 to a scalar flow; (ii) it does so on every complete graph with at least three vertices; and (iii) $h(x + y) = h(x) + h(y)$ for all $x, y \in \mathbb{R}$. If h is continuous at one point, these conditions are further equivalent to $h(t) = ct$ for some $c \in \mathbb{R}$. Consequently, among continuous odd transforms, linear maps are the only universal scalarity preservers. For a fixed collection of continuous maps $(h_r)_{r=1}^{R}$, the aggregate $A_{ij} = \sum_r w_r h_r(u_{ri} - u_{rj})$ is a score-difference flow for every collection of utilities and every $w \in \Delta_R$ if and only if every h_r is linear. The conclusion already applies at $R = 1$.*

This additive-scale result is the log-domain counterpart of consistency-preserving transformations for multiplicative pairwise-comparison matrices (Brunelli & Fedrizzi, 2015). The identity map gives $G_{\mathrm{src}}(\Theta) = 0 \Longrightarrow G_{\mathrm{dep}}(q;\Theta) = 0$. The expected-win choice $h(t) = \tanh(t/2)$ is nonlinear and breaks this universal implication. On a single-rubric triangle with utilities $(2, 1, 0)$, the transformed circulation is $2\tanh(1/2) - \tanh(1) > 0$. Scalar source logits can therefore induce a nonadditive deployment game under expected-win semantics while preserving every edge sign. For fixed utilities and weights, G_{dep} is the instance-specific diagnostic: a nonlinear map can produce either zero or positive residual. Source intervals govern projection error; the post-transform Hodge residual records the geometry presented to the decision rule.

Fix $\pi_0 \in \mathrm{ri}(\Delta_m)$, so $\pi_{0,i} > 0$ for every i. We choose the unique symmetric saddle policy of

$$\max_{p\in\Delta_m} \min_{s\in\Delta_m} p^{\top} A_q s - \tau\,\mathrm{KL}(p\|\pi_0) + \tau\,\mathrm{KL}(s\|\pi_0), \qquad \tau > 0. \tag{3.3}$$

This declared social-choice rule maps each skew payoff and reference policy to a unique mixed decision (Lemma B.2).

Proposition 3.2 (Scalar ranking as a special case). *If $A_{ij} = v_i - v_j$, the unique symmetric policy in (3.3) is $p_{A,i} = \frac{\pi_{0,i}\exp(v_i/\tau)}{\sum_k \pi_{0,k}\exp(v_k/\tau)}$. As $\tau \downarrow 0$, its mass concentrates on $\arg\max_i v_i$; the unregularized saddle policies are precisely the distributions supported on that set. Thus the Nash interface agrees with scalar ranking on additive payoffs and extends it to relational games.*

For a skew payoff A, define the unregularized exploitability of p as $\mathrm{Expl}_A(p) = \max_i e_i^{\top} Ap$.

Assumptions and falsifiability. Audit and recovery blocks target the same marginal cell probabilities, hence the same $\Theta^{\star}$. The analytic audit assumes bounded logits and independent within-cell votes, allowing worker-specific Bernoulli probabilities; rectangular inversion permits unbounded endpoints or intersection with a known bounded parameter box, while exact binomial intervals assume a common cell probability. Recovery uses independent sampled cells with conditional mean $\mathbb{E}(Y_t \mid X_t, E_t) = \sigma(\theta^{\star}_{X_t E_t})$, stated support or restricted curvature, and exact or approximate component ranks. The source-risk oracle post-processes one fixed full-flow estimator; its confidence selector uses an independent audit block and known or independently calibrated cyclic MSEs. Cross-fitting recycles observations while preserving this separation. Margins, rubrics, transforms, and deployment weights are fixed before testing. Table 1 records diagnostics and fallbacks.

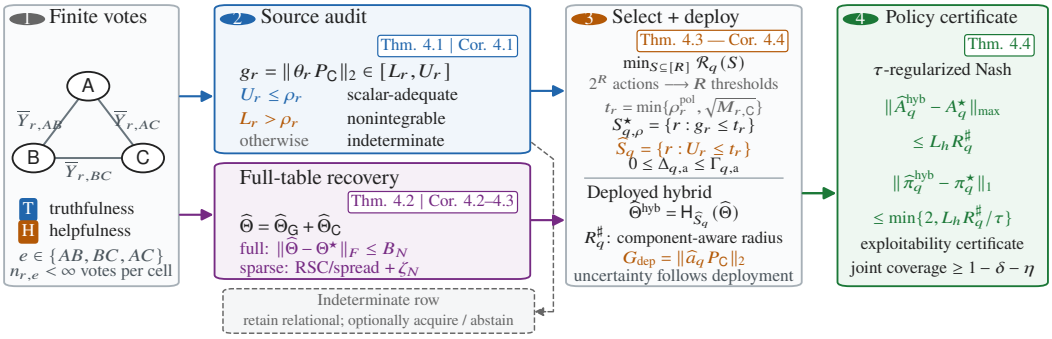


Figure 2: **Audit–recover–select–certify.** Parallel source audit (Theorem 4.1 and Corollary 4.1) and recovery (Theorem 4.2 and Corollaries 4.2 and 4.3) feed a finite post-fit source-risk program whose 2^R actions collapse to R comparisons (Theorem 4.3). Split-sample bounds control constrained regret (Corollary 4.4); the selected radius certifies policy and exploitability (Theorem 4.4). Indeterminate rows retain relational structure, with acquisition or abstention as operational options.

4 Test, Recover, Decide, Certify

This section audits safe itemwise representation, recovers the full table, then propagates both uncertainties to policy. The running example fixes prompt q_0, responses A, B, C, finite votes on every pair, truthfulness r_{tr}, helpfulness r_{hp}, and target $\Theta^\star$ (Figure 2).

4.1 A finite-sample scalar-adequacy audit

Index the $D = ZE$ context–edge cells by d. Within cell d, let $Y_{dk} \sim \text{Bernoulli}(p_{dk})$, $k \in [n_d]$, be independent with worker-specific probabilities, where $n_d \ge 1$. For fixed $a \in (0, 1/2)$, set $\bar{p}_d = n_d^{-1} \sum_k p_{dk} \in [a, 1-a]$, $\theta^\star_d = \text{logit}(\bar{p}_d)$, and $\bar{Y}_d = n_d^{-1} \sum_k Y_{dk}$. Dependence across cells is unrestricted. Hoeffding's inequality applies to these independent, non-identically distributed votes (Hoeffding, 1963). For $\delta \in (0,1)$, set $\widetilde{\theta}_d = \text{logit}\{\Pi_{[a/2, 1-a/2]}(\bar{Y}_d)\}$, $T = \left\|\widetilde{\Theta} P_{\mathsf{C}}\right\|_F$, $b^{\mathrm{H}}_{\boldsymbol{n},\delta} = L_a\{\frac{\log(2D/\delta)}{2} \sum_{d=1}^D n_d^{-1}\}^{1/2}$, and $L_a = \{(a/2)(1-a/2)\}^{-1}$.

Fix before observing outcomes an adequacy margin $\rho \ge 0$ for $G_{\text{src}}(\Theta^\star) = \|\Theta^\star P_{\mathsf{C}}\|_F$, and define the observable bracket $L^{\mathrm{H}}_{\boldsymbol{n},\delta} = (T - b^{\mathrm{H}}_{\boldsymbol{n},\delta})_+$ and $U^{\mathrm{H}}_{\boldsymbol{n},\delta} = T + b^{\mathrm{H}}_{\boldsymbol{n},\delta}$. The audit outputs *scalar-adequate* when $U^{\mathrm{H}}_{\boldsymbol{n},\delta} \le \rho$, *nonintegrable* when $L^{\mathrm{H}}_{\boldsymbol{n},\delta} > \rho$, and *indeterminate* otherwise. Setting $\rho = 0$ recovers exact scalarity testing; a nontrivial scalar-adequate branch requires positive resolution.

Theorem 4.1 (Worker-heterogeneous finite-sample scalar-adequacy audit)**.** *The bracket covers $G_{\text{src}}(\Theta^\star)$ with probability at least $1 - \delta$. In particular, the probability of either a scalar-adequate label when $G_{\text{src}}(\Theta^\star) > \rho$ or a nonintegrable label when $G_{\text{src}}(\Theta^\star) \le \rho$ is at most δ. This is unconditional control of an incorrect resolved outcome; separation from ρ governs resolution. The audit selects scalar-adequate with probability at least $1 - \delta$ when $G_{\text{src}}(\Theta^\star) \le \rho - 2b^{\mathrm{H}}_{\boldsymbol{n},\delta}$, and nonintegrable with the same probability when $G_{\text{src}}(\Theta^\star) > \rho + 2b^{\mathrm{H}}_{\boldsymbol{n},\delta}$. At $\rho = 0$, the nonintegrable branch is a level-δ scalarity test with power at least $1 - \delta$ beyond $2b^{\mathrm{H}}_{\boldsymbol{n},\delta}$.*

The fixed-direction Le Cam bound in Proposition B.3 identifies $n_{\text{eff}}(v)^{-1/2}$ as the unavoidable separation scale along a prespecified cycle direction.

Corollary 4.1 (Simultaneous adaptive representation)**.** *For $x = (q, r)$, let $T_x = \left\|\widetilde{\theta}_x P_{\mathsf{C}}\right\|_2$, $b_x = L_a \left\{\frac{\log(2D/\delta)}{2} \sum_e \frac{1}{n_{xe}}\right\}^{1/2}$. With probability at least $1 - \delta$, every row simultaneously satisfies $(T_x - b_x)_+ \le g_x(\Theta^\star) \le T_x + b_x$. Prespecified margins ρ_x support one family-wise rule: label row x scalar-adequate when $T_x + b_x \le \rho_x$, nonintegrable when $(T_x - b_x)_+ > \rho_x$, and indeterminate otherwise. Write $S = \{x : T_x + b_x \le \rho_x\}$; the illustrative resolved branch places (q_0, r_{tr}) in this projection-eligible set and (q_0, r_{hp}) outside it as nonintegrable.*

Both analytic and exact audits are instances of simultaneous confidence-region inversion. Let $[\ell_d, u_d]$ be any simultaneous probability rectangle with joint coverage at least $1-\delta$, using $[0,1]$ for a missing cell. Map it coordinatewise through the extended logit and form

$$\mathcal{C}_\delta = \{M : \operatorname{logit}(\ell_d) \le M_d \le \operatorname{logit}(u_d),\ d \in [D]\}. \tag{4.1}$$

Hoeffding and Clopper–Pearson intervals (Clopper & Pearson, 1934) give valid rectangles for heterogeneous-worker and common-probability cells. Proposition B.2 and Corollaries B.1 and B.2 converts them into rowwise brackets via distance to the gradient space and center–radius upper bounds. The same event certifies every row decision and predeclared cycle contrast; weak coverage of A–C drives its row toward indeterminate.

4.2 Structured full-table recovery

Fit the full table before the risk gate evaluates the projection-eligible truthfulness row. For observed votes $(X_t, E_t, Y_t)_{t=1}^N$, define $\mathcal{L}_N(M) = N^{-1}\sum_{t=1}^N [\log\{1+\exp(M_{X_tE_t})\} - Y_t M_{X_tE_t}]$ and estimate

$$\widehat{\Theta} \in \operatorname*{arg\,min}_{\|\Theta\|_{\max}\le\alpha} \{\mathcal{L}_N(\Theta) + \lambda_{\mathsf{G}} \|\Theta P_{\mathsf{G}}\|_* + \lambda_{\mathsf{C}} \|\Theta P_{\mathsf{C}}\|_*\}. \tag{4.2}$$

Separate penalties learn different gradient and cycle ranks. Section B.3 and Algorithm 1 gives the proximal map and workflow; the corollaries verify the theorem's score and curvature events.

Theorem 4.2 (Conditional Hodge low-rank oracle inequality)**.** *Let $\Delta = \widehat{\Theta} - \Theta^\star$. For $j \in \{\mathsf{G}, \mathsf{C}\}$, set $P_{\mathsf{G}} = P_{\mathsf{G}}$, $P_{\mathsf{C}} = P_{\mathsf{C}}$, $\Delta_j = \Delta P_j$, and let $\mathbb{T}_j$ be the nuclear-norm tangent space at $\Theta_j^\star = \Theta^\star P_j$. Assume $\|\Theta^\star\|_{\max} \le \alpha$, $\lambda_{\mathsf{G}}, \lambda_{\mathsf{C}} > 0$, the score bounds $\|\nabla\mathcal{L}_N(\Theta^\star)P_{\mathsf{G}}\|_{\mathrm{op}} \le \lambda_{\mathsf{G}}/2$, $\|\nabla\mathcal{L}_N(\Theta^\star)P_{\mathsf{C}}\|_{\mathrm{op}} \le \lambda_{\mathsf{C}}/2$ and restricted curvature with constant $\kappa > 0$ on the bounded weighted cone $\mathcal{K}_{\mathrm{exact}}(\alpha)$ in (B.6). If the component ranks are $(r_{\mathsf{G}}, r_{\mathsf{C}})$, define $C_\lambda = \{2\lambda_{\mathsf{G}}^2 r_{\mathsf{G}} + 2\lambda_{\mathsf{C}}^2 r_{\mathsf{C}}\}^{1/2}$. Then*

$$\|\Delta\|_F \le \frac{3C_\lambda}{\kappa}. \tag{4.3}$$

The supplement proves the claim and gives the tolerant cone and tail term for approximate rank.

Corollary 4.2 (Full-coverage random-design recovery)**.** *Under the i.i.d. full-support Bernoulli design, bounded-logit and component-rank conditions detailed in Section B.4, choose the matrix-Bernstein penalties $\lambda_j = 2u_j$ and put $q_\alpha = \sigma(\alpha)\{1-\sigma(\alpha)\}$. If $N\pi_{\min} \ge 8\log(2D/\eta)$, then, with probability at least $1-\eta$, $\left\|\widehat{\Theta} - \Theta^\star\right\|_F \le B_N := 6C_\lambda/(q_\alpha\pi_{\min})$.*

Corollary 4.3 (Sparse-design recovery with tolerance)**.** *Replace Theorem 4.2's restricted-curvature assumption by the realized Hodge-cone inequality in Section B.5, with constant $\kappa_{\mathrm{s}} > 0$ and tolerance ζ_N, and keep its other conditions. Then $\left\|\widehat{\Theta} - \Theta^\star\right\|_F / \sqrt{D} \le b_N^{\mathrm{sp}} := 3\sqrt{D}\, C_\lambda/(q_\alpha\kappa_{\mathrm{s}}) + \zeta_N/\sqrt{\kappa_{\mathrm{s}}}$. The radius $B_N = \sqrt{D}\, b_N^{\mathrm{sp}}$ feeds Theorem 4.4.*

The full conditions, sampling inequalities, and proofs of both specializations appear in Sections B.4 and B.5; coverage, tangent curvature, and nullity diagnose their design premises.

Theorem 4.3 (Exact fixed-estimator hybrid source-risk selection)**.** *Fix a query q, nonnegative weights w_{qr}, and one full-flow estimator $\widetilde{\theta}_{qr} = \theta_{qr}^\star + \varepsilon_{qr}$ with finite second moment under its sampling distribution. Write $M_{qr,j} = \mathbb{E}\|\varepsilon_{qr}P_j\|_2^2$ for $j \in \{\mathsf{G}, \mathsf{C}\}$. For a decision $z = (z_r)_{r=1}^R \in \mathcal{Z}_R := \{0,1\}^R$, let $z_r = 1$ project row r, let $z_r = 0$ retain it, and set $\widetilde{\theta}_{qr}^z = z_r\widetilde{\theta}_{qr}P_{\mathsf{G}} + (1-z_r)\widetilde{\theta}_{qr}$. Identify z with $S(z) = \{r : z_r = 1\}$ and write $\mathcal{R}_q(S) = \mathcal{R}_q(\mathbf{1}_S)$. The finite population-oracle problem is*

$$\begin{aligned} \mathcal{R}_q^\star &:= \min_{S\subseteq[R]} \mathcal{R}_q(S) = \min_{z\in\mathcal{Z}_R} \mathcal{R}_q(z), \\ \mathcal{R}_q(z) &:= \mathbb{E}\sum_r w_{qr}\|\widetilde{\theta}_{qr}^z - \theta_{qr}^\star\|_2^2 = \sum_r w_{qr}\left[M_{qr,\mathsf{G}} + z_r g_{qr}^2 + (1-z_r)M_{qr,\mathsf{C}}\right]. \end{aligned} \tag{4.4}$$

A canonical minimizer assigns risk ties to projection, with value

$$\begin{aligned} z_{qr}^\star &= \mathbf{1}\{g_{qr}^2 \le M_{qr,\mathsf{C}}\}, \quad z_q^\star \in \operatorname*{arg\,min}_{z\in\mathcal{Z}_R} \mathcal{R}_q(z), \quad S_q^\star = S(z_q^\star), \\ \mathcal{R}_q^\star &= \sum_r w_{qr}\left[M_{qr,\mathsf{G}} + \min\{g_{qr}^2, M_{qr,\mathsf{C}}\}\right]. \end{aligned} \tag{4.5}$$

The class $\mathcal{Z}_R$ contains 2^R actions. Hodge orthogonality removes the within-row gradient–cycle cross term; row-additive squared loss and the product action class remove cross-rubric coupling. Conditional on g_{qr} and $M_{qr,\mathsf{C}}$, the global program reduces to R threshold comparisons and evaluates in $O(R)$ time. Its exact gains over the two uniform actions are

$$\mathcal{R}_q(\varnothing)-\mathcal{R}_q(S_q^\star)=\sum_r w_{qr}(M_{qr,\mathsf{C}}-g_{qr}^2)_+,\quad \mathcal{R}_q([R])-\mathcal{R}_q(S_q^\star)=\sum_r w_{qr}(g_{qr}^2-M_{qr,\mathsf{C}})_+. \tag{4.6}$$

Thus the oracle strictly improves on both uniform representations when positive-weight rows occur on both strict sides of the threshold. If $\mathbb{E}\varepsilon_{qr}=0$ and $\mathrm{Cov}(\varepsilon_{qr}^\top)=\sigma_{qr}^2 I_E$, then $M_{qr,\mathsf{C}}=\sigma_{qr}^2\,\mathrm{rank}(P_{\mathsf{C}})$: retention spends one variance unit per cyclic degree of freedom.

Corollary 4.4 (Audit-calibrated constrained source-risk oracle)**.** *Let the audit data $\mathcal{D}_{\mathrm{a}}$ be independent of the fixed full-flow estimator, fix nonnegative policy-calibrated source-gap caps ρ_{qr}^{pol}, and suppose $M_{qr,\mathsf{C}}$ is known. Define*

$$\mathcal{Z}_{q,\rho}=\{z\in\{0,1\}^R : z_r g_{qr}\le z_r\rho_{qr}^{\mathrm{pol}}\ \forall r\},\qquad t_{qr}=\min\{\rho_{qr}^{\mathrm{pol}},\sqrt{M_{qr,\mathsf{C}}}\}. \tag{4.7}$$

A canonical constrained minimizer and its value are

$$\begin{aligned} &z_{qr,\rho}^\star=\mathbf{1}\{g_{qr}\le t_{qr}\},\qquad z_{q,\rho}^\star\in\operatorname*{arg\,min}_{z\in\mathcal{Z}_{q,\rho}}\mathcal{R}_q(z),\qquad S_{q,\rho}^\star=S(z_{q,\rho}^\star),\\ &\mathcal{R}_{q,\rho}^\star:=\min_{z\in\mathcal{Z}_{q,\rho}}\mathcal{R}_q(z)=\sum_r w_{qr}\left[M_{qr,\mathsf{G}}+M_{qr,\mathsf{C}}-\mathbf{1}\{g_{qr}\le\rho_{qr}^{\mathrm{pol}}\}(M_{qr,\mathsf{C}}-g_{qr}^2)_+\right]. \end{aligned} \tag{4.8}$$

Simultaneous brackets give $\widehat{z}_{qr}=\mathbf{1}\{U_{qr}\le t_{qr}\}$ and $\widehat{S}_q=S(\widehat{z}_q)$. With recovery-block risk, define $\Delta_{q,\mathrm{a}}=\mathcal{R}_q(\widehat{z}_q\mid\mathcal{D}_{\mathrm{a}})-\mathcal{R}_{q,\rho}^\star$. On any simultaneous-coverage event satisfying $0\le L_{qr}\le g_{qr}\le U_{qr}$ for every r,

$$0\le\Delta_{q,\mathrm{a}}=\sum_{r\in S_{q,\rho}^\star\setminus\widehat{S}_q}w_{qr}(M_{qr,\mathsf{C}}-g_{qr}^2)\le\Gamma_{q,\mathrm{a}}:=\sum_{r:L_{qr}\le t_{qr}<U_{qr}}w_{qr}(M_{qr,\mathsf{C}}-L_{qr}^2)_+. \tag{4.9}$$

No threshold-straddling interval implies exact constrained-oracle risk. If the maximal bracket width vanishes on simultaneous-coverage events whose probabilities tend to one, and no row lies on its threshold, then $\mathbb{P}(\widehat{S}_q=S_{q,\rho}^\star)\to 1$ and $\Gamma_{q,\mathrm{a}}\xrightarrow{\mathbb{P}}0$.

These results optimize source-logit risk for post-fit projections. Policy TV and exploitability may couple rubrics; Theorem 4.4 propagates source error through the fixed deployment rule. The calibration $L_h\{\sum_r w_{qr}(\rho_{qr}^{\mathrm{pol}})^2\}^{1/2}/\tau\le\varepsilon_{\pi,q}$ controls projection bias within an ℓ_1 policy budget, while the total certificate adds estimation error. The supplement treats unknown M_{C} and gives a minimax Gaussian benchmark (Proposition B.4).

4.3 The end-to-end audit-to-decision guarantee

The running branch projects truthfulness when $U_{q_0r_{\mathrm{tr}}}\le t_{q_0r_{\mathrm{tr}}}$ and retains relational helpfulness when $L_{q_0r_{\mathrm{hp}}}>\rho_{q_0r_{\mathrm{hp}}}^{\mathrm{pol}}$. Indeterminate rows default to relational retention; applications may acquire or abstain instead.

Theorem 4.4 (Finite-sample audit-to-decision guarantee)**.** *Suppose simultaneous rowwise brackets $[L_{qr},U_{qr}]$ cover $g_{qr}(\Theta^\star)\equiv g_{(q,r)}(\Theta^\star)$ with probability at least $1-\delta$. Fix source-gap caps ρ_{qr}^{pol} and thresholds $0\le t_{qr}\le\rho_{qr}^{\mathrm{pol}}$, using $t_{qr}=\min\{\rho_{qr}^{\mathrm{pol}},\sqrt{M_{qr,\mathsf{C}}}\}$ for Corollary 4.4. Set $\widehat{S}_q=\{r: U_{qr}\le t_{qr}\}$ for every query, project these rows, and retain the rest. The three-way labels at ρ_{qr}^{pol} call q resolved when every positive-weight row resolves; unresolved queries may instead trigger acquisition or abstention.*

Suppose also that, with probability at least $1-\eta$, recovery of the same $\Theta^\star$ supplies simultaneous weighted bounds

$$\sum_r w_{qr}\left\|\widehat{\theta}_{qr}-\theta_{qr}^\star\right\|_2^2\le B_{q,w}^2\qquad(q\in[Q]). \tag{4.10}$$

Let $S=\{(q,r): r\in\widehat{S}_q\}$ and define $(\mathsf{H}_S M)_{qr}=M_{qr}P_{\mathsf{G}}$ on S and M_{qr} elsewhere. Put $\widehat{\Theta}^{\mathrm{hyb}}=\mathsf{H}_S(\widehat{\Theta})$ and write $\Delta=\widehat{\Theta}-\Theta^\star$. If the same recovery event also supplies $\sum_r w_{qr}\|\Delta_{qr}P_{\mathsf{G}}\|_2^2\le$

$B^2_{q,\mathsf{G}}$ *and* $\|\Delta_{qr}P_{\mathsf{C}}\|_2 \le C_{qr}$*, set* $(R^\sharp_q)^2$ *to the minimum of* $B^2_{q,w} + \sum_{r\in\widehat{S}_q} w_{qr}U^2_{qr}$ *and* $B^2_{q,\mathsf{G}} + \sum_{r\notin\widehat{S}_q} w_{qr}C^2_{qr} + \sum_{r\in\widehat{S}_q} w_{qr}U^2_{qr}$*; otherwise use the first radius. Write* $\pi^\star_q = p_{A^\star_q}$ *and* $\widehat{\pi}^{\mathrm{hyb}}_q = p_{\widehat{A}^{\mathrm{hyb}}_q}$*. With probability at least* $1-\delta-\eta$*, every resolved row label is correct and every query satisfies*

$$\sum_r w_{qr}\left\|\widehat{\theta}^{\mathrm{hyb}}_{qr} - \theta^\star_{qr}\right\|_2^2 \le (R^\sharp_q)^2, \qquad \left\|\widehat{A}^{\mathrm{hyb}}_q - A^\star_q\right\|_{\max} \le L_h R^\sharp_q, \tag{4.11}$$

and its exact regularized policy satisfies

$$\left\|\widehat{\pi}^{\mathrm{hyb}}_q - \pi^\star_q\right\|_1 \le \min\{2, L_h R^\sharp_q/\tau\}, \qquad \mathrm{Expl}_{A^\star_q}(\widehat{\pi}^{\mathrm{hyb}}_q) \le L_h R^\sharp_q + \tau\log(1/\pi_{0,\min}). \tag{4.12}$$

A union bound gives the joint guarantee under arbitrary dependence. A global B_N *yields* $B_{q,w} = \sqrt{\max_r w_{qr}}B_N$*; optimization gap* $\varepsilon_{\mathrm{opt}}$ *adds to exploitability.*

For q_0, $\widehat{S}_{q_0} = \{r_{\mathrm{tr}}\}$ and $R^\sharp_{q_0} \le (B^2_{q_0,w} + w_{q_0 r_{\mathrm{tr}}}U^2_{q_0 r_{\mathrm{tr}}})^{1/2}$; component bounds remove projected cyclic estimation error. The condition $R^\sharp_q \le \tau\varepsilon_{\pi,q}/L_h$ certifies policy budget $\varepsilon_{\pi,q}$, while G_{dep} audits payoff additivity. The supplement proves the risk floor and local derivative certificate and gives Lean-checked identities (Propositions B.1 and B.5 and Sections B and D).

5 Experiments

Our experiments ask whether evidence follows the theory chain from representation uncertainty to downstream decision error. Controlled simulations test auditing, recovery, adaptive representation, and policy certification; ablations and sensitivity analyses identify the operative modules and the design boundary. Real-data studies then ask which representation conclusions the available evidence supports and how deployment transformations alter comparison geometry. Each result targets a theorem, assumption, or failure mode. Synthetic ground truth tests whether each guarantee activates at its stated boundary; real observations test whether the resulting diagnoses change practical representation choices.

Theory-linked design and metrics. Finite-cell coverage and resolution test the audit (Theorem 4.1 and Corollary 4.1); complete and sparse designs test recovery (Theorem 4.2 and Corollaries 4.2 and 4.3); paired selection tests source-risk optimality and regret (Theorem 4.3 and Corollary 4.4); payoff and policy errors test whether upstream guarantees survive deployment (Theorem 4.4). The registered design fixes six queries, four rubrics, four responses, independent audit and recovery blocks, and 96 paired vote-table realizations. Five baselines isolate the two fixed representations, the MSE gate, confidence regions, and Hodge decomposition; recovery grids add scalar and unstructured nuclear fits, taskwise BTL, weighted HodgeRank, independent cells, and pinned PCWST. Metrics follow the chain from source/logit error and KL/Brier loss to payoff error, policy TV, unsafe projection, oracle agreement, and resolution.

Finite-sample guarantees calibrate before comparison. Across 20,000 replications per condition, the three-way audit makes no wrong resolved decision under common-probability or heterogeneous-worker votes; the farthest registered nonintegrable condition is resolved nonintegrable at rate 1.000, while the margin boundary remains indeterminate. All 190,207 covered split-sample draws obey the constrained-regret envelope, with zero unsafe projections or remainder violations and zero covered-draw excess risk from $n = 1024$. All 768 above-threshold recovery fits satisfy B_N, with empirical/theoretical log–log slopes $-.488/-.505$, and a six-query audit records $0/300$ joint source, payoff, and policy-certificate failures at every budget (Table 3). Thus performance is compared only after the audit, recovery, and propagation guarantees are empirically calibrated.

Adaptive selection improves on both fixed extremes. Theorem 4.3 predicts a bias–variance crossover: project when cyclic estimation error exceeds projection bias and retain the relational representation otherwise. The risk-gated hybrid attains weighted source risk .6119 (95% CI .5898–.6339), 38.9% below forced scalar and 23.4% below forced relational, with zero observed unsafe projections, 100% registry-oracle agreement, and 76.2% resolution. Its maximum payoff error is lower than that of all five baselines (largest Holm-adjusted $p = 3.03\times10^{-21}$); policy TV is .0828 (Figure 3 and Table 4).

Ablations isolate the mechanism; sensitivity locates its boundary. Removing the MSE gate raises risk to .6923 and unsafe projections to 24.8%; plug-in selection gives risk .7901 with 7.6% unsafe

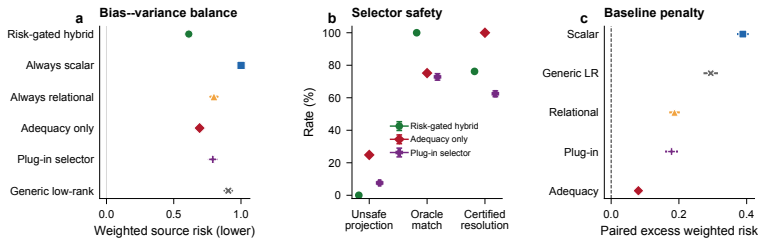


Figure 3: **Finite-sample representation selection follows the predicted bias–variance tradeoff. A**, the risk-gated hybrid has lower weighted source risk than both uniform representations. **B**, confidence-based selection records zero registry-oracle unsafe projections, 100% agreement, and 76.2% certified resolution. **C**, every fixed-choice or module-removal baseline has positive paired excess weighted risk. Together, the panels evaluate source-risk optimality, finite-sample selection, and downstream certification in Theorem 4.3, Corollary 4.4, and Theorem 4.4.

projections, while undecomposed low rank gives risk .9058. At 16 labels per cell, increasing cycle strength from zero to one moves scalar RMSE from .206 to .430 and Hodge RMSE from .244 to .312. The registered Hodge estimator achieves lower logit RMSE, KL, Brier, payoff RMSE, and policy TV than the pinned official PCWST solver with a shared-scale adapter in all nine cycle–budget cells (problem-level Holm $p < 1.7 \times 10^{-5}$). Unsafe projection remains zero across every registered allocation, margin, and cyclic-MSE perturbation; increasing audit labels from 128 to 8192 moves resolution from 0 to 94.3% and oracle agreement from 50% to 100% (Figure 8). Under structural missingness, truth-tangent curvature remains zero and nullity remains at least 36.8%, exposing the design boundary in Corollary 4.3 (Figures 5 and 10 and Table 5).

From controlled validation to real-world diagnosis. The simulations establish when the theoretical chain succeeds and where design geometry blocks recovery. The real-data studies expose two complementary representation phenomena. On Summarize-from-Feedback, 3,219 votes form 183 cells and 30 prompt graphs (Stiennon et al., 2020; OpenAI, 2020). At $\tau = .2$, exact boxes leave all 30 prompts indeterminate throughout $0 \leq \varepsilon_\pi \leq 2$, while heterogeneous-worker Hoeffding boxes are also indeterminate at $\rho = 0$; Holm-adjusted local scalarity tests reject none of the 30 graphs (Holm, 1979). Plug-in policy TV nevertheless reaches .360, and an acquire-on-indeterminate rule would request more evidence for every prompt. The operational conclusion is insufficient evidence for safe compression at the current annotation budget.

On 3,115 valid UltraFeedback records (Cui et al., 2024; OpenBMB, 2023), linear logits have zero cycle energy, although 82.8–87.1% of majority-payoff and 84.2–88.7% of expected-win matrices are nonadditive; median policy TV at $\tau = .2$ is .053–.057 and .017–.018, respectively. Nonadditivity remains 71.8–88.1% after removing any rubric and 32.6–72.7% for individual rubrics (Figure 14). This separates source nonintegrability from transformation-induced relational geometry: an exact scalar logit representation can still yield nonadditive deployment payoffs.

6 Discussion

Scalar and relational preference models typically fix the representation before testing whether finite evidence supports that choice for the downstream decision. We formulate representation as an auditable statistical action. Hodge geometry identifies the governing tradeoff: projection removes cyclic estimation variance while introducing cyclic approximation bias, so the source-risk optimum can select an itemwise or relational form for each rubric. Deployment adds a second insight: nonlinear payoff maps can recreate relational geometry even when every source logit is an exact score difference. The analysis fixes the full-flow estimator, rubric weights, deployment map, and aggregate population target. These scope choices define three research directions: joint representation–estimation, linked-rater dependence, and transformation-aware learning. Preference systems should therefore learn what to estimate, which representation the evidence supports, and how that choice shapes downstream decisions under finite-sample uncertainty and changing deployment semantics.

AI USE STATEMENT

The authors used generative AI tools to assist literature discovery, prose revision, experiment implementation and debugging, and translation of selected finite-dimensional proof statements into Lean. Every citation was checked against a primary source; every reported numerical result was regenerated from the archived code and data manifest; and the Lean kernel accepted files containing zero `sorry`, `admit`, unsafe definitions, or manuscript-specific axioms. The authors reviewed all AI-assisted work and take responsibility for the final text, claims, code, and artifacts.

ETHICS STATEMENT

The real-data studies use public releases and collect zero new human-subject data. The exact-cell audit uses released worker identifiers for within-cell deduplication; exported artifacts contain aggregate statistics and hashed prompt/candidate identifiers. UltraFeedback annotations were generated by GPT-4 and may inherit model, prompt-source, and rubric biases; their role here is an AI-feedback aggregation audit. The pairwise analyses target aggregate comparison geometry, while individual-level attribution requires linked within-rater designs. In deployment, a Nash policy is a normative choice rule whose rubrics, weights, and abstention conditions should be disclosed to affected users.

REPRODUCIBILITY STATEMENT

The supplement contains complete proofs, an assumption audit, data URLs and SHA-256 hashes, exclusion counts, deterministic seeds, solver diagnostics, and plotting code. Dependencies are pinned in the archived environment. The `LeanProofs` project includes its toolchain and mathlib revision, and its build output prints the axioms of every checked theorem. All empirical tables and figures are regenerated from machine-readable CSV or JSON outputs.

REFERENCES

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

Matteo Brunelli and Michele Fedrizzi. Axiomatic properties of inconsistency indices for pairwise comparisons. *Journal of the Operational Research Society*, 66(1):1–15, 2015. doi: 10.1057/jors.2013.135.

Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.

Shuo Chen and Thorsten Joachims. Modeling intransitivity in matchup and comparison data. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pp. 227–236, 2016. doi: 10.1145/2835776.2835787.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Banghua Zhu, Hao Zhang, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. Chatbot arena: An open platform for evaluating LLMs by human preference. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 8359–8388, 2024. URL https://proceedings.mlr.press/v235/chiang24b.html.

Paul F. Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4299–4307, 2017.

C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934. doi: 10.1093/biomet/26.4.404.

Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Bingxiang He, Wei Zhu, Yuan Ni, Guotong Xie, Ruobing Xie, Yankai Lin, Zhiyuan Liu, and Maosong Sun. ULTRAFEEDBACK: Boosting language models with scaled AI feedback. In *Proceedings of the 41st International Conference on*

Machine Learning, volume 235 of *Proceedings of Machine Learning Research*, pp. 9722–9744, 2024. URL https://proceedings.mlr.press/v235/cui24f.html.

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

Anisha Gunjal, Anthony Wang, Elaine Lau, Vaskar Nath, Yunzhong He, Bing Liu, and Sean Hendryx. Rubrics as rewards: Reinforcement learning beyond verifiable domains. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/cfd7bee7a651ee9af525098ef67a9e45-Abstract-Conference.html.

Helia Hashemi, Jason Eisner, Corby Rosset, Benjamin Van Durme, and Chris Kedzie. LLM-rubric: A multidimensional, calibrated approach to automated evaluation of natural language texts. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 13806–13834, Bangkok, Thailand, 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.745. URL https://aclanthology.org/2024.acl-long.745/.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. URL https://www.jstor.org/stable/4615733.

Yucong Huang, Xiucheng Li, Kaiqi Zhao, and Jing Li. Transitivity meets cyclicity: Explicit preference decomposition for dynamic large language model alignment, 2026. URL https://arxiv.org/abs/2605.17342. Accepted at the 43rd International Conference on Machine Learning.

Ruipeng Jia, Yunyi Yang, Wen Wang, Yuxin Wu, Yongbo Gai, Siyuan Tao, Mengyu Zhou, Jianhe Lin, Xiaoxi Jiang, and Guanjun Jiang. Open rubric system: Scaling reinforcement learning with pairwise adaptive rubric, 2026. URL https://arxiv.org/abs/2602.14069. Version 3.

Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1):203–244, 2011. doi: 10.1007/s10107-010-0419-x.

Seungone Kim, Jamin Shin, Yejin Cho, Joel Jang, Shayne Longpre, Hwaran Lee, Sangdoo Yun, Seongjin Shin, Sungdong Kim, James Thorne, and Minjoon Seo. Prometheus: Inducing fine-grained evaluation capability in language models. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/803485352e61e3ebf41221e4776c9fd4-Abstract-Conference.html.

Olga Klopp. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli*, 20(1): 282–303, 2014.

Vladimir Koltchinskii, Karim Lounici, and Alexandre B. Tsybakov. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Annals of Statistics*, 39(5):2302–2329, 2011.

Sze Ming Lee and Yunxiao Chen. Pairwise comparisons without stochastic transitivity: Model, theory and applications. *Journal of Machine Learning Research*, 27(161):1–49, 2026. URL https://www.jmlr.org/papers/v27/25-0217.html.

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

Jiachun Li, David Simchi-Levi, and Will Wei Sun. LLM evaluation as tensor completion: Low rank structure and semiparametric efficiency. *arXiv preprint arXiv:2604.05460*, 2026b. URL https://arxiv.org/abs/2604.05460.

Kaizhao Liu, Qi Long, Zhekun Shi, Weijie J. Su, and Jiancong Xiao. Statistical impossibility and possibility of aligning LLMs with human preferences: From Condorcet paradox to Nash equilibrium. *Annals of Statistics*, 54(4):2110–2135, 2026. doi: 10.1214/26-AOS2643.

LMArena. Arena human preference 55k. Hugging Face dataset, 2024. URL https://huggingface.co/datasets/lmarena-ai/arena-human-preference-55k. 57,477 pairwise battles; Apache-2.0 license.

R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

Anuran Makur and Japneet Singh. Minimax hypothesis testing for the Bradley–Terry–Luce model. *IEEE Transactions on Information Theory*, 71(12):9163–9202, 2025a. doi: 10.1109/TIT.2025.3605846.

Anuran Makur and Japneet Singh. Hypothesis testing for generalized Thurstone models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 42730–42764, 2025b. URL https://proceedings.mlr.press/v267/makur25a.html.

Roberto-Rafael Maura-Rivero, Marc Lanctot, Francesco Visin, and Kate Larson. Jackpot! alignment as a maximal lottery. *arXiv preprint arXiv:2501.19266*, 2025. URL https://arxiv.org/abs/2501.19266.

David C. McGarvey. A theorem on the construction of voting paradoxes. *Econometrica*, 21(4):608–610, 1953. doi: 10.2307/1907926.

Rémi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, Andrea Michi, Marco Selvi, Sertan Girgin, Nikola Momchev, Olivier Bachem, Daniel J. Mankowitz, Doina Precup, and Bilal Piot. Nash learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 36743–36768, 2024. URL https://proceedings.mlr.press/v235/munos24a.html.

Tianlong Nan, Xiaopeng Li, Christian Kroer, and Tianyi Lin. Efficient exploration for iterative Nash preference optimization, 2026. URL https://arxiv.org/abs/2606.01382.

Sahand Negahban and Martin J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Annals of Statistics*, 39(2):1069–1097, 2011.

Sahand Negahban and Martin J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, 13(53):1665–1697, 2012. URL https://www.jmlr.org/papers/v13/negahban12a.html.

Hisaya Okahara, Tomoyuki Nakagawa, and Shonosuke Sugasawa. The covariate-assisted bayesian intransitive Bradley–Terry model via combinatorial Hodge theory, 2026. URL https://arxiv.org/abs/2601.07158. Version 3.

OpenAI. Summarize from feedback: Human feedback data. Public dataset and code repository, 2020. URL https://github.com/openai/summarize-from-feedback. Feedback data under the repository's Modified MIT license; source Reddit TL;DR text retains separate CC BY 4.0 provenance.

OpenBMB. UltraFeedback dataset card. Hugging Face dataset, 2023. URL https://huggingface.co/datasets/openbmb/UltraFeedback/tree/40b436560ca83a8dba36114c22ab3c66e43f6d5e. Revision 40b436560ca83a8dba36114c22ab3c66e43f6d5e; MIT license.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pp. 27730–27744, 2022. doi: 10.52202/068431-2011.

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. In *Advances in Neural Information Processing Systems*, volume 36, pp. 53728–53741, 2023. doi: 10.52202/075280-2338.

Swastik Roy, Rajkumar Pujari, Tharindu Kumarage, Charith Peris, Rahul Gupta, Anna Rumshisky, Pradeep Natarajan, and Venkatesh Saligrama. PReMISE: Policy rubrics as measurement specifications for LLM judges, 2026. URL https://arxiv.org/abs/2605.30803.

Zhekun Shi, Kaizhao Liu, Qi Long, Weijie J. Su, and Jiancong Xiao. Fundamental limits of game-theoretic LLM alignment: Smith consistency and preference matching. *arXiv preprint arXiv:2505.20627*, 2025. URL https://arxiv.org/abs/2505.20627.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel M. Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F. Christiano. Learning to summarize from human feedback. In *Advances in Neural Information Processing Systems*, volume 33, pp. 3008–3021, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/1f89885d556929e98d3ef9b86448f951-Abstract.html.

Gokul Swamy, Christoph Dann, Rahul Kidambi, Steven Wu, and Alekh Agarwal. A minimaximalist approach to reinforcement learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 47345–47377, 2024. URL https://proceedings.mlr.press/v235/swamy24a.html.

Yue Wu, Zhiqing Sun, Huizhuo Yuan, Kaixuan Ji, Yiming Yang, and Quanquan Gu. Self-play preference optimization for language model alignment. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e48fa1c4f08fd1ae35d5df8352c3106d-Abstract-Conference.html.

Wenqian Xing. Attention limited reward learning, 2026. URL https://arxiv.org/abs/2607.04590.

Dehui Yang and Michael B. Wakin. Modeling and recovering non-transitive pairwise comparison matrices. In *2015 International Conference on Sampling Theory and Applications*, pp. 39–43, 2015. doi: 10.1109/SAMPTA.2015.7148846.

Haocheng Yang, Licheng Pan, Xiaoxi Li, Zhichao Chen, Zhiheng Zhang, Yuan Lu, Haoxuan Li, and Hao Wang. Rubrics on trial: Evolving rubrics from a single query via synthetic pairwise evidence. *arXiv preprint arXiv:2607.15092*, 2026. URL https://arxiv.org/abs/2607.15092.

Yifan Zhang, Ge Zhang, Yue Wu, Kangping Xu, and Quanquan Gu. Beyond Bradley–Terry models: A general preference model for language model alignment. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 76939–76965, 2025. URL https://proceedings.mlr.press/v267/zhang25de.html.

APPENDIX CONTENTS

Appendix	Contents	Page
Appendix A	**Assumptions, Diagnostics, and Fallbacks**	**15**
A.1	From matrix recovery to representation selection	15
A.2	Operational workflow	16
Appendix B	**Proofs**	**17**
B.1	Hodge representation and the exact scalar risk floor	17
B.2	Finite-sample scalarity testing	18
B.3	Conditional recovery of the two Hodge components	21
B.4	A design-specific recovery event	22
B.5	Sparse-design recovery and policy propagation	23
B.6	Payoff propagation, Nash stability, and exploitability	25
B.7	Local policy geometry	26
Appendix C	**Experiment Details and Additional Audits**	**28**
C.1	Theorem–evidence atlas	28
C.2	Direct finite-sample test check	29
C.3	Audit-driven multi-rubric hybrid	29
C.4	Split-sample audit-to-oracle calibration	31
C.5	End-to-end Bernoulli selector stress test	32
C.6	Complementary Hodge rank-mismatch phase diagram	32
C.7	Theory-matched Bernoulli simulation	34
C.8	Sparse-design curvature and tolerance diagnostic	36
C.9	Explicit oracle-bound calibration	37
C.10	Exact repeated-cell human comparison audit	37
C.11	Exploratory Arena aggregate audit	38
C.12	UltraFeedback processing and exclusions	39
C.13	Implementation safeguards	40
C.14	Claim–evidence map	40
Appendix D	**Lean 4 Verification Scope**	**42**

A Assumptions, Diagnostics, and Fallbacks

These conditions govern the stated audit, recovery, and policy guarantees for the specified population and design.

Condition	Role	Diagnostic	Fallback when violated
Bounded truth and fit $\|\Theta^\star\|_{\max} \le \alpha_0 < \alpha$, $\|\Theta\|_{\max} \le \alpha$	Keeps the logit map and likelihood curvature away from degeneracy on feasible segments.	Report edge win rates and active box constraints; vary (α_0, α).	Use one-sided probability intervals and partial identification for nearly deterministic edges.
Comparison support	Establishes restricted curvature for the observed design.	Plot prompt–rubric–edge counts and the smallest design eigenvalues on estimated tangent cones.	Restrict the target to covered edges, adaptively acquire missing comparisons, or abstain.
Low rank of ΘP_{G} and ΘP_{C}	Shares strength across prompt–rubric cells while retaining both Hodge components.	Singular-value scree plots, held-out Bernoulli likelihood, and rank sensitivity.	Replace exact rank by approximate-rank tails or fit cells independently when replication permits.
Fixed rubrics and weights	Separates measurement design from testing and policy choice.	Archive rubric text, preregister weights, and audit leave-one-rubric-out results.	Split data for rubric discovery and testing; report a robust policy set over plausible weights.
Prespecified adequacy margin ρ	Defines the largest source-cycle norm treated as scalar-adequate.	Report ρ on the same Frobenius scale as $G_{\mathrm{src}}(\Theta)$ and vary it through the policy-tolerance path.	Set $\rho = 0$ and disable the scalar-adequate branch when no positive tolerance is defensible.
Fixed post-fit estimator, calibrated M_{C}, and audit–recovery separation	Defines the source-risk action class and makes the confidence selector independent of recovery noise.	Archive estimator, folds, component-MSE calibration, and cross-fit assignments.	Use an independent calibration block, cross-fitting, or the conservative unknown-MSE rule in Section B.
Independent Bernoulli votes within each cell	Justifies the analytic threshold under heterogeneous probabilities; exact CP inversion additionally uses a common cell probability. Cross-cell independence is unrestricted.	Deduplicate worker–cell records; inspect worker load and acquisition clusters.	Use the heterogeneous Hoeffding audit, independent worker-level units, or a cluster-level threshold.
Positive temperature τ and full-support π_0	Gives a unique interior saddle and strong monotonicity.	Report $\pi_{0,\min}$ and a temperature path.	At $\tau = 0$, return a set of equilibria and use set-valued sensitivity analysis.

Table 1: Each assumption has a role, an observable diagnostic, and a prespecified fallback.

A.1 From matrix recovery to representation selection

The nesting is exact on the complete comparison graph. If centered taskwise scores form $U \in \mathbb{R}^{Z \times m}$ with $U\mathbf{1} = 0$, then the scalar edge table is

$$\Theta = UB^\top, \qquad \operatorname{rank}(\Theta) = \operatorname{rank}(U), \qquad \|\Theta\|_* = \sqrt{m}\,\|U\|_*. \tag{A.1}$$

Indeed, $B^\top B = mI - \mathbf{1}\mathbf{1}^\top$ for the complete graph, so $\Theta\Theta^\top = mUU^\top$. This identity locates low-rank BTL exactly inside the gradient branch and leaves the cyclic branch as a separate target.

Line of work	Scientific target	Observation, design, and structure	Primary guarantee / relation to this paper
Exact/noisy completion (Candès & Recht, 2009; Koltchinskii et al., 2011)	Recover a fixed low-rank real matrix.	Missing or additive-noise entries; incoherence or trace-norm structure; one ambient matrix.	Exact recovery or matrix-error rates; supplies low-rank estimation principles.
One-bit completion (Davenport et al., 2014)	Recover a fixed real matrix, or its linked observation law, from binary data.	Entrywise Bernoulli observations through $f(M_{ij})$; max-entry and low-rank/trace-norm control.	Frobenius or distributional error; a single penalty on Θ is the unstructured logistic analogue.
RSC and weighted/general sampling (Negahban & Wainwright, 2011; 2012; Klopp, 2014)	Quantify noisy recovery under nonuniform designs.	Decomposable nuclear regularization, spikiness, sampling RSC, or square-root nuclear estimation under general sampling.	Matrix/prediction error and minimax rates under paper-specific design conditions; supplies templates for our separately stated Hodge-cone RSC condition.
Scalar low-rank BTL (Li et al., 2026a;b)	Recover taskwise item scores under a fixed score-difference representation.	Pairwise Bernoulli votes with the additional constraint $\Theta P_{\mathsf{C}} = 0$.	The gradient-only branch; scalarity is assumed before estimation.
PCWST (Lee & Chen, 2026)	Recover a possibly intransitive comparison law within one response set.	A low-rank skew response-by-response matrix fitted from pairwise win counts; no cross-context Hodge sharing.	A modern recovery baseline in a general skew-matrix class containing Bradley–Terry, fitted without row-wise projection selection.
This paper	Select a rowwise representation, recover the full logit flow, and certify a decision.	$\Theta = \Theta P_{\mathsf{G}} + \Theta P_{\mathsf{C}}$; distinct component ranks and penalties; simultaneous positive-tolerance audit.	Recovery error plus representation bias, propagated to payoff, policy TV, and exploitability; unresolved rows retain the relational fit by default, with acquisition or abstention as operational options.

Table 2: Classical matrix completion supplies estimation tools inside our recovery lane; the paper's primary target is selective representation and its downstream decision cost.

A.2 Operational workflow

Algorithm 1 Adaptive representation audit and certified decision

1: Freeze rubrics, candidate sets, comparison graph, weights, error level δ, and adequacy margins.
2: Construct simultaneous rowwise brackets on an audit fold; fit $\widehat{\Theta}$ by (4.2) and calibrate M_{C} on independent recovery folds.
3: At ρ^{pol}, label each active row scalar-adequate, nonintegrable, or indeterminate.
4: Project rows with $U_{qr} \le t_{qr}$ and retain the rest, where $t_{qr} = \min\{\rho^{\mathrm{pol}}_{qr}, \sqrt{M_{qr,\mathsf{C}}}\}$; an indeterminate row defaults to relational retention.
5: Form $\widehat{A}_q$ from the selected hybrid table, compute $G_{\mathrm{dep}}(q)$, and solve (3.3).
6: Deploy the certified fallback, acquire comparisons, or abstain according to the prespecified operational rule; report source and deployment geometry, policy, projection sensitivity, exploitability radius, and diagnostics.

Theorem 4.4 combines audit and recovery events under arbitrary dependence. The source-risk regret statement in Corollary 4.4 uses independent folds or cross-fitting because its selected action is evaluated against recovery-block risk.

Incomplete comparison graphs. For the complete graph, the cycle space is $\ker(B^\top)$ and is spanned by triangle circulations. On a general observed graph, the orthogonal complement of $\mathrm{range}(B)$ also contains harmonic flow around unfilled holes. All projection identities below remain valid with the observed incidence matrix, but testing triangles alone is insufficient (Jiang et al., 2011).

Source geometry versus deployment geometry. A fixed linear combination of rubric scores remains a scalar score. Majority and nonlinear edgewise aggregation can break commutation with score differencing. Proposition 3.1 gives the sharp boundary: among continuous odd maps, universal preservation holds exactly for linear maps. Thus an aggregate Hodge residual can coexist with scalar-representable flow inside every rubric, even when $R = 1$. We report source and deployment residuals separately. The Lipschitz payoff-to-policy certificate in Theorem 4.4 covers identity and expected-win deployment; majority supplies a discontinuous geometry diagnostic. Individual-evaluator attribution calls for linked within-rater comparisons; McGarvey's theorem shows that individually transitive rankings can generate any majority tournament (McGarvey, 1953).

Source geometry	Deployment geometry	Interpretation
$G_{\mathrm{src}} = 0$	$G_{\mathrm{dep}} = 0$	both stages admit scalar representations
$G_{\mathrm{src}} = 0$	$G_{\mathrm{dep}} > 0$	transformation-induced payoff nonadditivity
$G_{\mathrm{src}} > 0$	$G_{\mathrm{dep}} = 0$	aggregation cancels or masks source circulation
$G_{\mathrm{src}} > 0$	$G_{\mathrm{dep}} > 0$	mixed source and transformation mechanisms

For the final row, nonlinearity prevents an additive attribution of the output cycle. The counterfactual residual $G_{\mathrm{dep}}(q; \Theta P_{\mathsf{G}})$ isolates whether the declared transform produces nonadditivity on the gradient projection itself.

B Proofs

B.1 Hodge representation and the exact scalar risk floor

Equip $\mathbb{R}^E$ with its Euclidean inner product. Since $\mathsf{G} = \mathrm{range}(B)$ and $\mathsf{C} = \ker(B^\top) = \mathsf{G}^\perp$, every edge flow f has a unique orthogonal decomposition $f = f_{\mathsf{G}} + f_{\mathsf{C}}$. Applying this row by row gives (3.1).

Lemma B.1 (Cycle criterion on a complete graph)**.** *For an antisymmetric edge flow f, the following are equivalent: (i) $f \in \mathsf{G}$; (ii) the circulation of f is zero on every closed walk; and (iii) the circulation is zero on every triangle.*

Proof. If $f = Bu$, then a walk sum telescopes, proving (i)⇒(ii), and (ii)⇒(iii) is immediate. For (iii)⇒(i), fix vertex 1, set $u_1 = 0$, and set $u_i = f_{i1}$ using the antisymmetric extension of the oriented edge vector. Zero circulation on triangle $(i, j, 1)$ gives $f_{ij} + f_{j1} + f_{1i} = 0$, hence $f_{ij} = u_i - u_j$. Thus $f = Bu$. □

Proof of Proposition 3.1. Assume preservation on K_3. Fix $x, y \in \mathbb{R}$ and choose vertex utilities $u_1 = x + y$, $u_2 = y$, and $u_3 = 0$. On edges $(1, 2), (2, 3), (1, 3)$ the scalar flow is $(x, y, x + y)$. By Lemma B.1, its transformed circulation vanishes:

$$h(x) + h(y) - h(x + y) = 0,$$

so h is additive. Conversely, additivity gives $h(u_i - u_j) = h(u_i) - h(u_j)$, which is a scalar flow on every complete graph. This proves the first three-way equivalence and also shows that preservation itself forces $h(0) = 0$ and oddness. If h is continuous at one point, additivity makes it continuous everywhere; $h(q) = qh(1)$ for rational q, and density extends this identity to all real arguments. Thus $h(t) = ct$. Every such linear map is additive, completing the regular case.

For the aggregate statement, linear active maps give $A_{ij} = \sum_r w_r c_r (u_{ri} - u_{rj}) = v_i - v_j$ with $v_i = \sum_r w_r c_r u_{ri}$. Conversely, place all weight on an active rubric. Universal aggregate preservation invokes the single-map equivalence and forces its transform to be linear. The same argument establishes the $R = 1$ special case. □

Proof of Proposition 3.2. Write $A = v\mathbf{1}^\top - \mathbf{1}v^\top$, so for probability vectors p, s,

$$p^\top A s = p^\top v - v^\top s.$$

The regularized saddle objective therefore separates into a function of p and the negative of the same function of s. Put

$$Z_\tau = \sum_{j=1}^{m} \pi_{0,j} \exp(v_j/\tau), \qquad p_i^\tau = \frac{\pi_{0,i} \exp(v_i/\tau)}{Z_\tau}.$$

For every $p \in \Delta_m$, direct substitution gives the Gibbs identity

$$p^\top v - \tau \,\mathrm{KL}(p\|\pi_0) = \tau \log Z_\tau - \tau \,\mathrm{KL}(p\|p^\tau).$$

Hence p^τ is the unique maximizer; the separated minimization in s has the same unique solution. Thus the symmetric saddle policy is $p^\tau = s^\tau$ with the displayed coordinates.

Let $v_{\max} = \max_i v_i$ and $I^\star = \{i : v_i = v_{\max}\}$. Dividing numerator and denominator by $\exp(v_{\max}/\tau)$ shows that, as $\tau \downarrow 0$, the mass outside $I^\star$ vanishes and

$$p_i^\tau \longrightarrow \frac{\pi_{0,i}\mathbf{1}\{i \in I^\star\}}{\sum_{j \in I^\star} \pi_{0,j}}.$$

In particular, a unique top-scoring response receives limiting mass one; with ties, the limit is the reference policy restricted and renormalized on the top-score set. At $\tau = 0$, the payoff remains $p^\top v - v^\top s$. A row strategy is a best response exactly when it maximizes $p^\top v$, and a column strategy is a best response exactly when it maximizes $v^\top s$. Each condition is equivalent to support on $I^\star$, which proves the stated characterization of all unregularized saddle policies. □

Proposition B.1 (Exact scalar approximation floor)**.** *For every edge-logit table $\Theta^\star$,*

$$\inf_{MP_{\mathsf{C}}=0} \|\Theta^\star - M\|_F^2 = \|\Theta^\star P_{\mathsf{C}}\|_F^2 . \tag{B.1}$$

Proof. Write $\Theta^\star = \Theta^\star P_{\mathsf{G}} + \Theta^\star P_{\mathsf{C}}$. For any scalar table M satisfying $MP_{\mathsf{C}} = 0$,

$$\begin{aligned}\|\Theta^\star - M\|_F^2 &= \|(\Theta^\star P_{\mathsf{G}} - M) + \Theta^\star P_{\mathsf{C}}\|_F^2 \\ &= \|\Theta^\star P_{\mathsf{G}} - M\|_F^2 + \|\Theta^\star P_{\mathsf{C}}\|_F^2 ,\end{aligned}$$

because the two summands occupy orthogonal column subspaces. The lower bound is attained at $M = \Theta^\star P_{\mathsf{G}}$. □

For a fixed target table and Euclidean loss, this is the irreducible scalar-approximation floor. A change in conditioning variables, population, rubric definition, or edge weights defines a new projection target.

B.2 Finite-sample scalarity testing

Proof of the upper bounds in Theorem 4.1. Let $\bar{Y}_d = n_d^{-1}\sum_{k=1}^{n_d} Y_{dk}$, $\bar{p}_d = n_d^{-1}\sum_{k=1}^{n_d} p_{dk}$, and $\varepsilon_d = \sqrt{\log(2D/\delta)/(2n_d)}$. Cellwise Hoeffding inequalities and a union bound give the event

$$\mathcal{A}_{\boldsymbol{n},\delta} = \bigcap_{d=1}^{D} \{|\bar{Y}_d - \bar{p}_d| \le \varepsilon_d\}, \qquad \mathbb{P}(\mathcal{A}_{\boldsymbol{n},\delta}) \ge 1 - \delta. \tag{B.2}$$

Hoeffding's inequality requires independence and boundedness within a cell; the success probabilities may differ across workers. The union bound permits arbitrary dependence between cells. Projection onto an interval is nonexpansive and $\bar{p}_d \in [a, 1-a] \subset [a/2, 1-a/2]$. The derivative of $\mathrm{logit}(p)$ is $1/\{p(1-p)\}$, bounded by L_a on the clipping interval. Therefore, on $\mathcal{A}_{\boldsymbol{n},\delta}$,

$$\left\|\widetilde{\Theta} - \Theta^\star\right\|_F \le L_a \left(\sum_{d=1}^{D} \varepsilon_d^2\right)^{1/2} = b^{\mathrm{H}}_{\boldsymbol{n},\delta}. \tag{B.3}$$

Orthogonal projection is contractive, so $\left\|(\widetilde{\Theta}-\Theta^\star)P_{\mathsf{C}}\right\|_F \le b^{\mathrm{H}}_{\boldsymbol{n},\delta}$ on the same event. The triangle and reverse-triangle inequalities then give

$$(T - b^{\mathrm{H}}_{\boldsymbol{n},\delta})_+ \le G_{\mathrm{src}}(\Theta^\star) \le T + b^{\mathrm{H}}_{\boldsymbol{n},\delta},$$

which proves bracket coverage and rules out either resolved error on an event of probability at least $1-\delta$. If $G_{\mathrm{src}}(\Theta^\star) \le \rho - 2b^{\mathrm{H}}_{\boldsymbol{n},\delta}$, then $T + b^{\mathrm{H}}_{\boldsymbol{n},\delta} \le \rho$ on this event. If $G_{\mathrm{src}}(\Theta^\star) > \rho + 2b^{\mathrm{H}}_{\boldsymbol{n},\delta}$, then $T - b^{\mathrm{H}}_{\boldsymbol{n},\delta} > \rho$. These implications prove the two separated-regime guarantees. At $\rho = 0$, the lower endpoint is positive exactly when $T > b^{\mathrm{H}}_{\boldsymbol{n},\delta}$, which recovers the stated size and power bounds for ϕ^{H}_δ. □

Proof of Corollary 4.1. On the event $\mathcal{A}_{\boldsymbol{n},\delta}$ in (B.2), the coordinatewise argument above gives, simultaneously for every row x,

$$\left\|\widetilde{\theta}_x - \theta^\star_x\right\|_2 \le L_a \left(\sum_e \varepsilon^2_{xe}\right)^{1/2} = b_x.$$

Contractivity of P_{C} and the reverse-triangle inequality therefore imply

$$|T_x - g_x(\Theta^\star)| \le \left\|(\widetilde{\theta}_x - \theta^\star_x)P_{\mathsf{C}}\right\|_2 \le b_x,$$

which proves all rowwise brackets on one event of probability at least $1-\delta$. A row selected by its upper endpoint satisfies $g_x(\Theta^\star) \le T_x + b_x \le \rho_x$, and a row selected by its lower endpoint satisfies $g_x(\Theta^\star) > \rho_x$. □

Proposition B.2 (Generic rectangular confidence-region inversion)**.** *Let*

$$\mathcal{C}_\delta = \{M : M^-_d \le M_d \le M^+_d,\ d \in [D]\}$$

be any possibly unbounded random rectangle satisfying $\mathbb{P}_\Theta(\Theta \in \mathcal{C}_\delta) \ge 1-\delta$. Define

$$L^{\square}_\delta = \inf_{M\in\mathcal{C}_\delta} \|MP_{\mathsf{C}}\|_F, \qquad L^{\square}_{\delta,x} = \inf_{m\in\mathcal{C}_{\delta,x}} \|mP_{\mathsf{C}}\|_2,$$

where $\mathcal{C}_{\delta,x}$ is the row-x rectangle. If all global endpoints are finite, set

$$M^0 = \frac{M^- + M^+}{2}, \qquad r_\delta = \frac{1}{2}\left\|M^+ - M^-\right\|_F, \qquad \overline{U}^{\square}_\delta = \left\|M^0 P_{\mathsf{C}}\right\|_F + r_\delta; \tag{B.4}$$

otherwise set $\overline{U}^{\square}_\delta = \infty$. Rowwise, use

$$M^0_x = \frac{M^-_x + M^+_x}{2}, \qquad r_{\delta,x} = \frac{1}{2}\left\|M^+_x - M^-_x\right\|_2, \qquad \overline{U}^{\square}_{\delta,x} = \left\|M^0_x P_{\mathsf{C}}\right\|_2 + r_{\delta,x}$$

when that row has finite endpoints, and $\overline{U}^{\square}_{\delta,x} = \infty$ otherwise. Then, with probability at least $1-\delta$, simultaneously

$$G_{\mathrm{src}}(\Theta) \in [L^{\square}_\delta, \overline{U}^{\square}_\delta] \quad \text{and} \quad g_x(\Theta) \in [L^{\square}_{\delta,x}, \overline{U}^{\square}_{\delta,x}] \quad \text{for every } x.$$

Consequently, global or rowwise three-way rules based on prespecified margins have resolved-error probability at most δ. At zero margin, $L^{\square}_{\delta,x} > 0$ if and only if $\mathcal{C}_{\delta,x} \cap \mathrm{range}(B) = \varnothing$; testing the complementary event is a linear feasibility problem. For any predeclared contrasts $c_1, \dots, c_K$, the intervals

$$I_k = \left[\inf_{M\in\mathcal{C}_\delta} \langle c_k, \mathrm{vec}(M)\rangle,\ \sup_{M\in\mathcal{C}_\delta} \langle c_k, \mathrm{vec}(M)\rangle\right]$$

simultaneously cover their population values with probability at least $1-\delta$.

Proof. Work on the joint-coverage event $\Theta \in \mathcal{C}_\delta$. The definitions of the infima immediately give $L^{\square}_\delta \le G_{\mathrm{src}}(\Theta)$ and $L^{\square}_{\delta,x} \le g_x(\Theta)$ for every row. If the global endpoints are finite, every $M \in \mathcal{C}_\delta$ satisfies $\left\|M - M^0\right\|_F \le r_\delta$. Contractivity of P_{C} and the triangle inequality yield

$$\|MP_{\mathsf{C}}\|_F \le \left\|M^0 P_{\mathsf{C}}\right\|_F + \left\|(M - M^0)P_{\mathsf{C}}\right\|_F \le \overline{U}^{\square}_\delta.$$

The same argument in one row gives all rowwise upper bounds. The infinite-endpoint convention is automatically safe. These conclusions hold on one event of probability at least $1-\delta$, proving the simultaneous brackets and the resolved-error statement.

For a row rectangle, the condition $mP_{\mathsf{C}} = 0$ is equivalent to $m \in \mathrm{range}(B)$. Hence intersection gives $L^{\square}_{\delta,x} = 0$. Conversely, the linear image of the polyhedral set $\mathcal{C}_{\delta,x}$ under $m \mapsto mP_{\mathsf{C}}$ is a closed polyhedron; if its distance from zero is zero, it contains zero. Thus a zero lower endpoint is equivalent to feasibility. In the probability-box construction (4.1), this feasibility problem is

$$\mathrm{logit}(\ell_{xe}) \le (Bu_x)_e \le \mathrm{logit}(u_{xe}) \quad \text{for every observed edge } e,$$

with an arbitrary centering constraint on u_x. Finally, on the same joint-coverage event, every population contrast lies between its infimum and supremum over $\mathcal{C}_\delta$, proving simultaneous contrast coverage. □

Corollary B.1 (Heterogeneous-worker Hoeffding rectangle)**.** *Under the sampling conditions of Theorem 4.1, let*

$$\varepsilon_d = \sqrt{\frac{\log(2D/\delta)}{2n_d}}, \qquad \ell_d^{\mathrm{H}} = (\bar{Y}_d - \varepsilon_d)_+, \qquad u_d^{\mathrm{H}} = \min\{1, \bar{Y}_d + \varepsilon_d\}.$$

Mapping these intervals through the extended logit in (4.1) gives a rectangle with joint coverage at least $1-\delta$ for the population-average logits $\theta_d = \mathrm{logit}(\bar{p}_d)$. This conclusion requires independence and boundedness within each cell, permits worker-specific success probabilities, and permits arbitrary dependence across cells. Under the additional bounded-logit condition in Theorem 4.1, the finite logit rectangle

$$\mathcal{C}_\delta^{\mathrm{H,fin}} = \left\{ M : \left| M_d - \widetilde{\theta}_d \right| \le L_a \varepsilon_d, d \in [D] \right\}$$

also has joint coverage at least $1-\delta$. Applying Proposition B.2 to either rectangle gives simultaneous global and rowwise brackets; for the finite rectangle, its center–radius upper endpoints equal the analytic upper endpoints in Theorem 4.1 and Corollary 4.1, while optimizing the lower endpoint can only sharpen the reverse-triangle lower bound.

Proof. Hoeffding's inequality for sums of independent bounded variables gives $\mathbb{P}\{\left|\bar{Y}_d - \bar{p}_d\right| > \varepsilon_d\} \le \delta/D$ for each cell even when the Bernoulli probabilities differ. A union bound establishes joint probability-interval coverage, and monotonicity of the extended logit gives the first logit rectangle. On the same event, clipping is nonexpansive and the logit is L_a-Lipschitz on $[a/2, 1-a/2]$, so $\left|\widetilde{\theta}_d - \theta_d\right| \le L_a \varepsilon_d$ simultaneously. This proves coverage of the finite rectangle. Its global Euclidean radius is $b^{\mathrm{H}}_{\boldsymbol{n},\delta}$ and its row-x radius is b_x, which proves the final claims by Proposition B.2. □

Corollary B.2 (Exact iid-binomial rectangle)**.** *Suppose $Y_{dk} \overset{\mathrm{iid}}{\sim} \mathrm{Bernoulli}(p_d)$ within each cell. Give each two-sided Clopper–Pearson interval $[\ell_d^{\mathrm{E}}, u_d^{\mathrm{E}}]$ marginal error δ/D and map its endpoints through the extended logit as in (4.1). The resulting rectangle has joint coverage at least $1-\delta$ under arbitrary cross-cell dependence. It therefore satisfies every conclusion of Proposition B.2. If a logit endpoint is infinite, the corresponding safe upper endpoint is infinite; a known bounded-logit parameter box may be intersected before applying the proposition.*

Proof. Exact binomial coverage gives $\mathbb{P}\{p_d \in [\ell_d^{\mathrm{E}}, u_d^{\mathrm{E}}]\} \ge 1 - \delta/D$ for every cell (Clopper & Pearson, 1934). The union bound gives simultaneous coverage, and monotonicity of the extended logit places Θ in the resulting rectangle. Proposition B.2 completes the proof. □

Clopper–Pearson coverage targets a common binomial probability. Under worker heterogeneity, Corollary B.1 targets the finite-worker population average $\bar{p}_d$.

Proposition B.3 (Fixed-direction audit lower bound)**.** *For a unit cycle direction v in row x, let $n_{\mathrm{eff}}(v) = \sum_e n_{xe} v_e^2$. Along the common-probability submodel $\Theta_t = t e_x v^\top$, every measurable test ϕ of $H_0 : \Theta = 0$ against $H_1 : \Theta = \Theta_t$ satisfies $P_0(\phi = 1) + P_{\Theta_t}(\phi = 0) \ge 3/4$ whenever*

$$|t| \le \min\left\{ n_{\mathrm{eff}}(v)^{-1/2}, \frac{\log\{(1-a)/a\}}{\|v\|_\infty} \right\}.$$

Proof. It suffices to restrict the parameter space to a one-dimensional submodel. Fix $x_0 \in [Z]$, let $v \in \mathsf{C}$ have $\|v\|_2 = 1$, set $\Theta_t = te_{x_0}v^\top$, and assume

$$|t| \le \min\left\{ n_{\mathrm{eff}}(v)^{-1/2}, \frac{\log\{(1-a)/a\}}{\|v\|_\infty} \right\}.$$

Compare $H_0 : \Theta = 0$ with $H_1 : \Theta = \Theta_t$, observing $n_{x_0 e}$ independent Bernoulli votes on edge e in row x_0. The second bound ensures that every alternative probability remains in $[a, 1-a]$. For $p_s = \sigma(s)$,

$$k(s) := \mathrm{KL}\{\mathrm{Bern}(p_s) \| \,\mathrm{Bern}(1/2)\}$$

is even, $k(0) = 0$, and for $s \ge 0$ satisfies $k'(s) = sp_s(1-p_s) \le s/4$. Hence $k(s) \le s^2/8$ for all s. Tensorization gives

$$\mathrm{KL}(P_{\Theta_t} \| P_0) \le \frac{t^2}{8} \sum_e n_{x_0 e} v_e^2 = \frac{t^2 n_{\mathrm{eff}}(v)}{8}.$$

Pinsker's inequality implies $\mathrm{TV}(P_{\Theta_t}, P_0) \le |t|\sqrt{n_{\mathrm{eff}}(v)}/4$. For $|t| \le n_{\mathrm{eff}}(v)^{-1/2}$ this is at most $1/4$. Le Cam's two-point identity then gives, for every test ϕ,

$$P_0(\phi = 1) + P_{\Theta_t}(\phi = 0) \ge 1 - \mathrm{TV}(P_{\Theta_t}, P_0) \ge \frac{3}{4}.$$

Thus the direction-specific separation scale is $n_{\mathrm{eff}}(v)^{-1/2}$. Growing-dimensional detection adds nuisance directions and has its own minimax boundary. □

B.3 Conditional recovery of the two Hodge components

Exact constrained proximal map. For $R(\Theta) = \lambda_{\mathsf{G}} \|\Theta P_{\mathsf{G}}\|_* + \lambda_{\mathsf{C}} \|\Theta P_{\mathsf{C}}\|_*$, orthogonality gives

$$\mathrm{prox}_{\eta R}(X) = \mathrm{SVT}_{\eta\lambda_{\mathsf{G}}}(XP_{\mathsf{G}})P_{\mathsf{G}} + \mathrm{SVT}_{\eta\lambda_{\mathsf{C}}}(XP_{\mathsf{C}})P_{\mathsf{C}}.$$

The box projection is entrywise clipping. Dykstra's algorithm alternates these exact proximal maps with correction variables and computes the proximal map of $\eta R + I_{\{\|\Theta\|_{\max} \le \alpha\}}$ used by the solver.

For a rank-r matrix $M = U\Sigma V^\top$, let $\mathbb{T}(M) = \{UX^\top + YV^\top : X, Y \text{ conformable}\}$ be the tangent space and $P_{\mathbb{T}(M)}$ its Frobenius projection. We use the standard nuclear-norm decomposability inequality (Negahban & Wainwright, 2011)

$$\|M\|_* - \|M + H\|_* \le \left\|P_{\mathbb{T}(M)}H\right\|_* - \left\|P^{\perp}_{\mathbb{T}(M)}H\right\|_*. \tag{B.5}$$

With T and O defined in the proof below, set

$$\mathcal{K}_{\mathrm{exact}}(\alpha) = \{\Delta : \mathsf{O}(\Delta) \le 3\mathsf{T}(\Delta), \quad \|\Theta^\star + \Delta\|_{\max} \le \alpha\}. \tag{B.6}$$

Restricted curvature on this bounded cone means

$$\mathcal{L}_N(\Theta^\star + \Delta) - \mathcal{L}_N(\Theta^\star) - \langle \nabla\mathcal{L}_N(\Theta^\star), \Delta\rangle \ge \frac{\kappa}{2}\|\Delta\|_F^2 \quad \text{for all } \Delta \in \mathcal{K}_{\mathrm{exact}}(\alpha). \tag{B.7}$$

Proof of Theorem 4.2. Write $\Delta_{\mathsf{G}} = \Delta P_{\mathsf{G}}$ and $\Delta_{\mathsf{C}} = \Delta P_{\mathsf{C}}$, and abbreviate

$$\mathsf{T}(\Delta) = \lambda_{\mathsf{G}}\|P_{\mathbb{T}_{\mathsf{G}}}\Delta_{\mathsf{G}}\|_* + \lambda_{\mathsf{C}}\|P_{\mathbb{T}_{\mathsf{C}}}\Delta_{\mathsf{C}}\|_*,$$
$$\mathsf{O}(\Delta) = \lambda_{\mathsf{G}}\left\|P_{\mathbb{T}_{\mathsf{G}}^\perp}\Delta_{\mathsf{G}}\right\|_* + \lambda_{\mathsf{C}}\left\|P_{\mathbb{T}_{\mathsf{C}}^\perp}\Delta_{\mathsf{C}}\right\|_*.$$

Optimality, convexity of $\mathcal{L}_N$, the score bounds, and decomposability give, before any curvature assumption is invoked,

$$0 \ge \mathcal{L}_N(\Theta^\star + \Delta) - \mathcal{L}_N(\Theta^\star) + R(\Theta^\star + \Delta) - R(\Theta^\star) \ge \frac{1}{2}\mathsf{O}(\Delta) - \frac{3}{2}\mathsf{T}(\Delta), \tag{B.8}$$

where R is the penalty in (4.2). Feasibility gives $\|\Theta^\star + \Delta\|_{\max} \le \alpha$; hence $\Delta \in \mathcal{K}_{\mathrm{exact}}(\alpha)$ and (B.7) applies. Reusing the same score and decomposability bounds yields

$$\frac{\kappa}{2}\|\Delta\|_F^2 \le \frac{3}{2}\mathsf{T}(\Delta) - \frac{1}{2}\mathsf{O}(\Delta) \le \frac{3}{2}\mathsf{T}(\Delta)$$

$$\leq \frac{3}{2} C_\lambda \left(\|\Delta_{\mathsf{G}}\|_F^2 + \|\Delta_{\mathsf{C}}\|_F^2 \right)^{1/2} = \frac{3}{2} C_\lambda \|\Delta\|_F .$$

The last line uses $\operatorname{rank}(P_{\mathbb{T}(M)} H) \leq 2\operatorname{rank}(M)$, $\|H\|_* \leq \sqrt{\operatorname{rank}(H)}\,\|H\|_F$, Cauchy–Schwarz across the two components, and $\Delta_{\mathsf{G}} \perp \Delta_{\mathsf{C}}$. Dividing by $\|\Delta\|_F$ proves (4.3); the zero-error case is immediate.

For approximate rank, let $M_j = \Theta^\star P_j$, let M_{j,r_j} be its best rank-r_j truncation, redefine $\mathbb{T}_j = \mathbb{T}(M_{j,r_j})$ and T, O accordingly, and set

$$\mathfrak{t} = \lambda_{\mathsf{G}} \|M_{\mathsf{G}} - M_{\mathsf{G},r_{\mathsf{G}}}\|_* + \lambda_{\mathsf{C}} \|M_{\mathsf{C}} - M_{\mathsf{C},r_{\mathsf{C}}}\|_* , \qquad C_\lambda = \{2\lambda_{\mathsf{G}}^2 r_{\mathsf{G}} + 2\lambda_{\mathsf{C}}^2 r_{\mathsf{C}}\}^{1/2}.$$

For each component, approximate decomposability reads

$$\|M_j\|_* - \|M_j + H\|_* \leq \left\|P_{\mathbb{T}_j} H\right\|_* - \left\|P_{\mathbb{T}_j^\perp} H\right\|_* + 2\left\|M_j - M_{j,r_j}\right\|_* .$$

Summing with weights gives the tolerant cone $\mathsf{O}(\Delta) \leq 3\mathsf{T}(\Delta) + 4\mathfrak{t}$. If restricted curvature holds on its intersection with $\{\|\Theta^\star + \Delta\|_{\max} \leq \alpha\}$, then

$$\|\Delta\|_F \leq \frac{3C_\lambda + \sqrt{9C_\lambda^2 + 16\kappa\mathfrak{t}}}{2\kappa} \leq \frac{3C_\lambda}{\kappa} + 2\sqrt{\frac{\mathfrak{t}}{\kappa}}.$$

□

B.4 A DESIGN-SPECIFIC RECOVERY EVENT

Write $\pi_{\cdot e} = \sum_x \pi_{xe}$. For $j \in \{\mathsf{G}, \mathsf{C}\}$, define the explicit design radii

$$\begin{aligned}
\ell_j &= \max_e \|P_j e_e\|_2 , \qquad t_\eta = \log \frac{4(Z+E)}{\eta}, \\
v_j &= \frac{1}{4} \max\left\{ \max_x \sum_e \pi_{xe} \|P_j e_e\|_2^2 , \|P_j \operatorname{diag}(\pi_{\cdot 1}, \ldots, \pi_{\cdot E}) P_j\|_{\mathrm{op}} \right\}, \\
u_j &= \sqrt{\frac{2v_j t_\eta}{N}} + \frac{2\ell_j t_\eta}{3N}, \qquad \lambda_j = 2u_j.
\end{aligned} \tag{B.9}$$

Detailed conditions for Corollary 4.2. Let $(X_t, E_t)_{t=1}^N$ be i.i.d. with probabilities $\pi_{xe} \geq \pi_{\min} > 0$; conditional on these indices, let the Y_t be independent Bernoulli variables with means $\sigma(\theta^\star_{X_t E_t})$. Assume $\eta \in (0,1)$, $\|\Theta^\star\|_{\max} \leq \alpha$, and prespecified rank bounds $\operatorname{rank}(\Theta^\star P_j) \leq r_j$ for $j \in \{\mathsf{G}, \mathsf{C}\}$. Put $q_\alpha = \sigma(\alpha)\{1 - \sigma(\alpha)\}$ and choose $\lambda_j = 2u_j$ in (B.9). If $N\pi_{\min} \geq 8\log(2D/\eta)$, then, with probability at least $1 - \eta$,

$$\left\|\widehat{\Theta} - \Theta^\star\right\|_F \leq B_N := \frac{6C_\lambda}{q_\alpha \pi_{\min}}. \tag{B.10}$$

For the running slice, full coverage samples all six rubric–edge cells; rectangular matrix Bernstein, multinomial coverage, and logistic curvature yield B_N. Missing edges instead require spread or an explicit tolerance beyond rank and boundedness (Davenport et al., 2014; Negahban & Wainwright, 2011; Klopp, 2014).

Proof of Corollary 4.2. Write $\xi_t = \sigma(\theta^\star_{X_t E_t}) - Y_t$ and

$$G_j = \nabla \mathcal{L}_N(\Theta^\star) P_j = \frac{1}{N} \sum_{t=1}^N \xi_t e_{X_t} (P_j e_{E_t})^\top .$$

Conditional centering gives $\mathbb{E}(\xi_t \mid X_t, E_t) = 0$ and $\mathbb{E}(\xi_t^2 \mid X_t, E_t) \leq 1/4$. Each unscaled summand has operator norm at most ℓ_j. Its two matrix-variance proxies are bounded by

$$\frac{1}{4} \max_x \sum_e \pi_{xe} \|P_j e_e\|_2^2 , \qquad \frac{1}{4} \|P_j \operatorname{diag}(\pi_{\cdot 1}, \ldots, \pi_{\cdot E}) P_j\|_{\mathrm{op}} .$$

Rectangular matrix Bernstein therefore gives

$$\mathbb{P}\{\|G_j\|_{\mathrm{op}} > u_j\} \le (Z+E)e^{-t_\eta} = \eta/4.$$

A union bound over $j \in \{\mathsf{G}, \mathsf{C}\}$ establishes both score inequalities with probability at least $1-\eta/2$.

Let $N_{xe} = \sum_{t=1}^{N} \mathbf{1}\{(X_t, E_t) = (x, e)\}$. Multiplicative Chernoff gives

$$\mathbb{P}\{N_{xe} < N\pi_{xe}/2\} \le e^{-N\pi_{xe}/8}.$$

The assumed sample size and a union bound imply, with probability at least $1-\eta/2$, that $N_{xe}/N \ge \pi_{\min}/2$ for all cells. On this event,

$$Q_N(\Delta) = \frac{1}{N}\sum_{t=1}^{N} \Delta^2_{X_t E_t} = \sum_{x,e} \frac{N_{xe}}{N}\Delta^2_{xe} \ge \frac{\pi_{\min}}{2}\|\Delta\|_F^2 .$$

Along the feasible segment the logistic second derivative is at least q_α, so its Bregman divergence is at least $(q_\alpha/2)Q_N(\Delta)$. Hence (B.7) holds for every feasible Δ with $\kappa = q_\alpha\pi_{\min}/2$. Substituting $\lambda_j = 2u_j$ and this κ into Theorem 4.2 proves (B.10). □

For uniform sampling, $\pi_{\min} = 1/D$ and, with $d_j = \mathrm{rank}(P_j)$,

$$v_j \le \frac{1}{4}\max\{d_j/D, 1/E\}.$$

The resulting normalized squared-error rate, up to universal constants and the Bernstein second-order term, is

$$\frac{\left\|\widehat{\Theta} - \Theta^\star\right\|_F^2}{D} \lesssim \frac{\{r_{\mathsf{G}}\max(Z, d_{\mathsf{G}}) + r_{\mathsf{C}}\max(Z, d_{\mathsf{C}})\}\log\{(Z+E)/\eta\}}{q_\alpha^2 N}.$$

An unobserved concentrated direction precludes uniform positive curvature over the unrestricted boxed low-rank class: a matrix supported on one unobserved cell has both projected component ranks at most one and is observationally indistinguishable from zero. Sparse recovery therefore restricts such directions through incoherence or spread, or retains a sampling-dependent tolerance.

B.5 Sparse-design recovery and policy propagation

Detailed conditions for Corollary 4.3. Let $D = ZE$, $q_\alpha = \sigma(\alpha)\{1-\sigma(\alpha)\}$, $C_\lambda = \{2\lambda_{\mathsf{G}}^2 r_{\mathsf{G}} + 2\lambda_{\mathsf{C}}^2 r_{\mathsf{C}}\}^{1/2}$, and $Q_N(\Delta) = N^{-1}\sum_{t=1}^{N}\Delta^2_{X_tE_t}$. Assume every condition of Theorem 4.2 except its restricted-curvature inequality, and impose on the same weighted cone

$$Q_N(\Delta) \ge \frac{\kappa_{\mathrm{s}}}{D}\|\Delta\|_F^2 - \zeta_N^2, \qquad \kappa_{\mathrm{s}} > 0. \tag{B.11}$$

Then

$$\frac{\left\|\widehat{\Theta} - \Theta^\star\right\|_F}{\sqrt{D}} \le b_N^{\mathrm{sp}} := \frac{3\sqrt{D}}{q_\alpha\kappa_{\mathrm{s}}}C_\lambda + \frac{\zeta_N}{\sqrt{\kappa_{\mathrm{s}}}}. \tag{B.12}$$

Here $B_N = \sqrt{D}\, b_N^{\mathrm{sp}}$ feeds Theorem 4.4. The condition requires curvature on the Hodge error cone; matrix-completion spread conditions provide templates for verifying it under sparse sampling (Negahban & Wainwright, 2012; Davenport et al., 2014).

Proof of Corollary 4.3. The score bounds and decomposability argument in the proof of Theorem 4.2 give both $\Delta \in \mathcal{K}_{\mathrm{exact}}(\alpha)$ and the basic inequality

$$D_{\mathcal{L}_N}(\Theta^\star + \Delta, \Theta^\star) \le \frac{3}{2}C_\lambda\|\Delta\|_F . \tag{B.13}$$

Along the feasible segment, logistic curvature is at least q_α, and therefore $D_{\mathcal{L}_N}(\Theta^\star+\Delta, \Theta^\star) \ge (q_\alpha/2)Q_N(\Delta)$. Combining this inequality with (B.11) yields, for $x = \|\Delta\|_F$,

$$\frac{q_\alpha\kappa_{\mathrm{s}}}{D}x^2 - 3C_\lambda x - q_\alpha\zeta_N^2 \le 0.$$

The positive root and $\sqrt{a+b} \le \sqrt{a}+\sqrt{b}$ give $x/\sqrt{D} \le 3\sqrt{D}C_\lambda/(q_\alpha\kappa_{\mathrm{s}}) + \zeta_N/\sqrt{\kappa_{\mathrm{s}}}$, proving (B.12). The policy claims follow from Lemmas B.3 to B.5. □

Corollary 4.3 conditions on the realized sampling operator through (B.11). When $N < D$, spiky directions can remain unobserved under uniform entry sampling. Matrix-completion results verify restricted curvature under explicit incoherence or spikiness conditions with a sampling-dependent tolerance (Negahban & Wainwright, 2012). The empirical design reports realized coverage and null directions below.

Proof of Theorem 4.3. For an action with $z_r = 1$, orthogonality of the Hodge subspaces gives the pointwise identity

$$\|\widetilde{\theta}_{qr}P_{\mathsf{G}} - \theta^{\star}_{qr}\|_2^2 = \|\varepsilon_{qr}P_{\mathsf{G}} - \theta^{\star}_{qr}P_{\mathsf{C}}\|_2^2 = \|\varepsilon_{qr}P_{\mathsf{G}}\|_2^2 + g_{qr}^2.$$

For $z_r = 0$, the retained-row error satisfies $\|\varepsilon_{qr}\|_2^2 = \|\varepsilon_{qr}P_{\mathsf{G}}\|_2^2 + \|\varepsilon_{qr}P_{\mathsf{C}}\|_2^2$. Taking expectations and summing with the nonnegative weights proves the risk identity in (4.4); the cross term vanishes pointwise, so unbiasedness is unnecessary. Rearranging gives

$$\mathcal{R}_q(z) = \sum_r w_{qr}(M_{qr,\mathsf{G}} + M_{qr,\mathsf{C}}) + \sum_r w_{qr}z_r(g_{qr}^2 - M_{qr,\mathsf{C}}).$$

The feasible set is a Cartesian product and the objective has no cross-rubric term. The displayed canonical action sets $z_r = 1$ when $g_{qr}^2 \le M_{qr,\mathsf{C}}$ and therefore attains the smaller coordinate cost, proving (4.5). Positive-weight rows away from equality have a unique action; risk ties and zero-weight rows can have additional minimizers. Conditional on g_{qr} and $M_{qr,\mathsf{C}}$, this coordinatewise solution evaluates the global optimum in $O(R)$ operations without enumerating 2^R actions. Subtracting the optimum from either uniform sum gives (4.6). A positive-weight row with $g_{qr}^2 > M_{qr,\mathsf{C}}$ makes all-scalar risk strictly larger, and one with $g_{qr}^2 < M_{qr,\mathsf{C}}$ makes all-relational risk strictly larger. Finally, mean-zero isotropic error gives $M_{qr,\mathsf{C}} = \operatorname{tr}(P_{\mathsf{C}}\sigma_{qr}^2 I P_{\mathsf{C}}) = \sigma_{qr}^2 \operatorname{tr}(P_{\mathsf{C}}) = \sigma_{qr}^2 \operatorname{rank}(P_{\mathsf{C}})$. □

Proof of Corollary 4.4. For any row with $z_r = 1$, feasibility in (4.7) requires $g_{qr} \le \rho_{qr}^{\mathrm{pol}}$; among feasible rows, the canonical action projects when $g_{qr}^2 \le M_{qr,\mathsf{C}}$. Their intersection is $g_{qr} \le t_{qr}$. The coordinate cost is therefore g_{qr}^2 on this event and $M_{qr,\mathsf{C}}$ otherwise, proving both the minimizer and optimal value in (4.8). Condition on the audit data. Independence makes $\widehat{S}_q$ fixed relative to the recovery block, so (4.4) applies conditionally with the same component MSEs. On simultaneous bracket coverage, $r \in \widehat{S}_q$ implies $g_{qr} \le U_{qr} \le t_{qr}$; hence $\widehat{S}_q \subseteq S^{\star}_{q,\rho}$. The risks differ only on rows projected by the constrained oracle and retained by the audit rule, giving the equality in (4.9). Every such row obeys $L_{qr} \le g_{qr} \le t_{qr} < U_{qr}$ and $0 \le M_{qr,\mathsf{C}} - g_{qr}^2 \le (M_{qr,\mathsf{C}} - L_{qr}^2)_+$. Summation proves the bound. If no interval straddles t_{qr}, every row in $S^{\star}_{q,\rho}$ has $U_{qr} \le t_{qr}$, so the selectors coincide. For the asymptotic claim, finite R and the absence of threshold ties give $d_q = \min_r |g_{qr} - t_{qr}| > 0$. On a simultaneous-coverage event with $\max_r(U_{qr} - L_{qr}) < d_q$, a row with $g_{qr} < t_{qr}$ has $U_{qr} < t_{qr}$, while a row with $g_{qr} > t_{qr}$ has $L_{qr} > t_{qr}$; hence no interval straddles its threshold, $\widehat{S}_q = S^{\star}_{q,\rho}$, and $\Gamma_{q,\mathrm{a}} = 0$. Coverage probability tending to one and maximal width converging to zero on those events prove selector consistency and $\Gamma_{q,\mathrm{a}} \xrightarrow{\mathbb{P}} 0$. □

Unknown component MSE. Use an independent calibration block and intersect bracket coverage for g_{qr} with simultaneous coverage $M^{-}_{qr,\mathsf{C}} \le M_{qr,\mathsf{C}} \le M^{+}_{qr,\mathsf{C}}$. A projected row satisfies $g_{qr} \le \rho_{qr}^{\mathrm{pol}}$ and $g_{qr}^2 \le U_{qr}^2 \le M^{-}_{qr,\mathsf{C}} \le M_{qr,\mathsf{C}}$. A retained row satisfies either $g_{qr} > \rho_{qr}^{\mathrm{pol}}$ or $g_{qr}^2 \ge L_{qr}^2 > M^{+}_{qr,\mathsf{C}} \ge M_{qr,\mathsf{C}}$. Hence the rule that projects under $U_{qr} \le \rho_{qr}^{\mathrm{pol}}$ and $U_{qr}^2 \le M^{-}_{qr,\mathsf{C}}$, retains under $L_{qr} > \rho_{qr}^{\mathrm{pol}}$ or $L_{qr}^2 > M^{+}_{qr,\mathsf{C}}$, and otherwise abstains agrees with the constrained oracle on every resolved row.

Proposition B.4 (Gaussian sequence minimax benchmark)**.** *Let orthogonal spaces* G *and* C *have positive dimensions* d_{G} *and* d_{C}*, let* $a \ge 0$ *and* $\sigma > 0$*, and observe* $Y_j = \theta_j + \sigma Z_j$ *for* $j \in \{\mathsf{G}, \mathsf{C}\}$*, where the* Z_j *are independent standard Gaussian vectors,* $\theta_{\mathsf{G}} \in \mathbb{R}^{d_{\mathsf{G}}}$*, and* $\|\theta_{\mathsf{C}}\|_2 \le a$*. Under squared error,*

$$d_{\mathsf{G}}\sigma^2 + \Phi(-1)\min\{a^2, d_{\mathsf{C}}\sigma^2\} \le \inf_{\widehat{\theta}} \sup_{\theta_{\mathsf{G}},\, \|\theta_{\mathsf{C}}\|_2 \le a} \mathbb{E}\|\widehat{\theta} - \theta\|_2^2 \le d_{\mathsf{G}}\sigma^2 + \min\{a^2, d_{\mathsf{C}}\sigma^2\}, \quad \text{(B.14)}$$

where Φ *is the standard normal distribution function.*

Proof. For the upper bound, estimate the unrestricted gradient component by Y_{G} and estimate the cycle component by zero when $a^2 \le d_{\mathsf{C}}\sigma^2$, or by Y_{C} otherwise. The component risks add and give the stated upper bound.

When $a = 0$, the cycle term vanishes and the unrestricted Gaussian-location lower bound for the gradient component is exact. For $a > 0$, use the product prior $\theta_{\mathsf{G}} \sim N(0, T^2 I)$ and let the d_{C} cycle coordinates be independent uniform signs times $b = \min\{a/\sqrt{d_{\mathsf{C}}}, \sigma\}$. The prior lies in the parameter set, and posterior factorization makes its Bayes risk additive. The gradient Bayes risk is $d_{\mathsf{G}}\sigma^2 T^2/(\sigma^2 + T^2)$. For one cycle coordinate, an estimate with the wrong sign incurs squared loss at least b^2, while the optimal sign test between $N(-b, \sigma^2)$ and $N(b, \sigma^2)$ errs with probability $\Phi(-b/\sigma) \ge \Phi(-1)$. Its Bayes risk is therefore at least $b^2\Phi(-1)$, and summing gives $\Phi(-1)\min\{a^2, d_{\mathsf{C}}\sigma^2\}$. The minimax risk dominates every Bayes risk; letting $T \to \infty$ proves the lower bound. This proposition targets Gaussian sequence estimation; Theorem 4.2 covers Bernoulli low-rank recovery. □

Optimality hierarchy. Equation (4.6) gives dominance over the two uniform representations. Theorem 4.3 solves the finite weighted source-logit risk problem over all 2^R rubric-wise actions: Hodge orthogonality removes each gradient–cycle cross term, row-additive loss removes cross-rubric objective terms, and the product action class yields the $O(R)$ closed form once oracle inputs are available. Corollary 4.4 confines finite-sample regret to threshold-straddling rows and recovers the constrained action as brackets shrink away from the boundary. These results optimize post-fit source-logit risk for a fixed estimator. Proposition B.4 gives an all-estimator minimax benchmark for its Gaussian cycle-ball class. Theorem 4.4 supplies policy and exploitability certificates from selected source error; the optimization claim remains the source-risk result.

The global all-query certificate can be sharpened for the experiment's mean-policy metric. Define $e_q^2 = \sum_{r,e} w_{qr}\Delta_{qre}^2$ and $w_{\max} = \max_{q,r} w_{qr}$. Cauchy–Schwarz and Lipschitzness give $\left\|\widehat{A}_q - A_q^\star\right\|_{\max} \le L_h e_q$ and $\sum_q e_q^2 \le w_{\max}\|\Delta\|_F^2$. Hence

$$\frac{1}{Q}\sum_{q=1}^{Q} \mathrm{TV}(\widehat{\pi}_q, \pi_q^\star) \le \min\left\{1, \frac{L_h\|\Delta\|_F}{2\tau}\sqrt{\frac{w_{\max}}{Q}}\right\}, \tag{B.15}$$

$$\frac{1}{Q}\sum_{q=1}^{Q} \mathrm{Expl}_{A_q^\star}(\widehat{\pi}_q) \le L_h\|\Delta\|_F\sqrt{\frac{w_{\max}}{Q}} + \tau\log(1/\pi_{0,\min}). \tag{B.16}$$

For uniform rubric weights, (B.15) becomes $Q^{-1}\sum_q \mathrm{TV}(\widehat{\pi}_q, \pi_q^\star) \le \min\{1, L_h\sqrt{E}\, b_N^{\mathrm{sp}}/(2\tau)\}$.

B.6 Payoff propagation, Nash stability, and exploitability

Lemma B.2 (Well-posed entropy-regularized skew saddle)**.** *Let $A = -A^\top$, $\tau > 0$, and $\pi_0 \in \mathrm{ri}(\Delta_m)$. The game in (3.3) has a unique saddle pair (p_A, s_A), both components lie in $\mathrm{ri}(\Delta_m)$, and $p_A = s_A$.*

Proof. With $0\log 0 = 0$, the objective is continuous on the compact product simplex, concave in p, and convex in s, so Sion's theorem gives a saddle pair. For fixed s, strict concavity and the Lagrange equations give the unique Gibbs response

$$p_i(s) = \frac{\pi_{0,i}\exp\{(As)_i/\tau\}}{\sum_j \pi_{0,j}\exp\{(As)_j/\tau\}},$$

and the minimizing Gibbs response to fixed p has the analogous form with $-A^\top p$; full support of π_0 makes both responses interior. Every saddle pair is therefore interior.

On the product simplex define

$$F_A(p, s) = \begin{bmatrix} -As + \tau\{\log(p/\pi_0) + \mathbf{1}\} \\ A^\top p + \tau\{\log(s/\pi_0) + \mathbf{1}\} \end{bmatrix}.$$

If $z = (p, s)$ and $z' = (p', s')$ are saddle pairs, their variational inequalities imply $\langle F_A(z) - F_A(z'), z - z'\rangle \le 0$. Direct expansion cancels the bilinear terms and gives

$$\langle F_A(z) - F_A(z'), z - z'\rangle = \tau\{\mathrm{J}(p, p') + \mathrm{J}(s, s')\} \ge 0.$$

Thus both Jeffreys divergences vanish and the saddle pair is unique. Finally, skew symmetry gives $L_A(s,p) = -L_A(p,s)$ for the saddle objective L_A; the saddle inequalities therefore show that (s_A, p_A) is also a saddle pair. Uniqueness yields $p_A = s_A$. □

Lemma B.3 (Payoff Lipschitzness). *If the deployment weights are fixed, nonnegative, and sum to one, then*

$$\left\|\widehat{A}_q - A^\star_q\right\|_{\max} \le L_h \max_{r,e} |\widehat{\theta}_{qre} - \theta^\star_{qre}| \le L_h \left\|\widehat{\Theta} - \Theta^\star\right\|_F .$$

Proof. Apply the declared Lipschitz inequality inside (3.2), use nonnegative weights summing to one, and bound the maximum entry by the Frobenius norm. Identity deployment has $L_h = 1$; expected-win deployment has $L_h = 1/2$ because $h'(t) = \frac{1}{2}\operatorname{sech}^2(t/2)$. □

Lemma B.4 (Regularized saddle stability). *Let $z_A = (p_A, s_A)$ and $z_{\widetilde{A}} = (p_{\widetilde{A}}, s_{\widetilde{A}})$ be the regularized saddle points with common (τ, π_0). If $\left\|A - \widetilde{A}\right\|_{\max} \le \delta$, then*

$$\tau\{\mathrm{J}(p_A, p_{\widetilde{A}}) + \mathrm{J}(s_A, s_{\widetilde{A}})\} \le 4\delta, \tag{B.17}$$

where $\mathrm{J}(p,q) = \mathrm{KL}(p\|q) + \mathrm{KL}(q\|p)$. If both games are skew-symmetric, $p_A = s_A$, $p_{\widetilde{A}} = s_{\widetilde{A}}$, and $\left\|p_A - p_{\widetilde{A}}\right\|_1 \le \min\{2, \delta/\tau\}$.

Proof. On the product simplex define the saddle operator

$$F_A(p,s) = \begin{bmatrix} -As + \tau\{\log(p/\pi_0) + \mathbf{1}\} \\ A^\top p + \tau\{\log(s/\pi_0) + \mathbf{1}\} \end{bmatrix}.$$

For $z = (p,s)$ and $z' = (p',s')$, bilinear cross-terms cancel and

$$\langle F_A(z) - F_A(z'), z - z' \rangle = \tau\{\mathrm{J}(p,p') + \mathrm{J}(s,s')\}.$$

The two saddle points solve their variational inequalities. Add those inequalities and insert $F_A(z_{\widetilde{A}})$. The remainder contains two inner products, each bounded by 2δ, because multiplying a max-norm bounded matrix by a probability vector has infinity norm at most δ and two probability vectors are at ℓ_1 distance at most two. This proves (B.17).

For skew games, uniqueness and symmetry give $p_A = s_A$ and $p_{\widetilde{A}} = s_{\widetilde{A}}$. Put $D = A - \widetilde{A}$ and $d = p_A - p_{\widetilde{A}}$. Adding the two row-optimality variational inequalities and using $d^\top A d = 0$ gives the sharper relation

$$2\tau\, \mathrm{J}(p_A, p_{\widetilde{A}}) \le 2\left|d^\top D p_{\widetilde{A}}\right| \le 2\delta \|d\|_1 .$$

Pinsker's inequality in both directions gives $\mathrm{J}(p_A, p_{\widetilde{A}}) \ge \|d\|_1^2$. Canceling a nonzero $\|d\|_1$ and using the diameter of the simplex proves the stated minimum; the zero-distance case is immediate. □

B.7 Local policy geometry

Proposition B.5 (Computable local policy certificate). *Let $V \in \mathbb{R}^{m\times(m-1)}$ satisfy $V^\top V = I$ and $V^\top \mathbf{1} = 0$, so its columns form an orthonormal basis of the simplex tangent space. Let p_A be the regularized symmetric policy. The matrix $J_A = V^\top\{\tau \operatorname{diag}(p_A)^{-1} - A\}V$ is invertible and, for every skew H,*

$$\left\|J_A^{-1}\right\|_{\mathrm{op}} \le 1/\tau, \qquad Dp_A[H] = V J_A^{-1} V^\top H p_A. \tag{B.18}$$

The induced ℓ_∞-to-ℓ_1 norm $\Lambda(A, p_A) = \sup_{H^\top = -H,\, \|H\|_{\max}\le 1} \left\|V J_A^{-1} V^\top H p_A\right\|_1$ is computable; the proof below also gives a quadratic remainder and its validity radius.

Proof of Proposition B.5. The interior symmetric policy satisfies the tangent-space KKT equation

$$V^\top\{A p_A - \tau \log(p_A/\pi_0)\} = 0. \tag{B.19}$$

The additive derivative of the KL term disappears because $V^\top \mathbf{1} = 0$. The symmetric part of J_A is $\tau V^\top \operatorname{diag}(p_A)^{-1} V \succeq \tau I$; the A term is skew-symmetric. Hence $\|J_A z\|_2 \|z\|_2 \ge z^\top J_A z \ge \tau \|z\|_2^2$, which proves invertibility and $\left\|J_A^{-1}\right\|_{\mathrm{op}} \le 1/\tau$.

Differentiate (B.19) along $A + tH$ and write the tangent derivative as $V\dot{z}$. At $t = 0$,

$$-J_A \dot{z} + V^\top H p_A = 0,$$

which yields (B.18) by the implicit-function theorem. The displayed optimization defining Λ is over a compact polytope. Since the objective is convex in the independent upper-triangular entries of H, its maximum is attained at a sign vertex; exact enumeration is therefore available when the action set is small.

For a finite perturbation, let $A = \widehat{A} + H$, let $\widehat{p} = p_{\widehat{A}}$, $h = p_A - \widehat{p}$, $\delta = \|H\|_{\max}$, and $\widehat{\gamma} = \left\|J_{\widehat{A}}^{-1}\right\|_{\mathrm{op}}$. If $\delta \le \tau \widehat{p}_{\min}/2$, the global bound in Lemma B.4 gives $\|h\|_2 \le \|h\|_1 \le \delta/\tau \le \widehat{p}_{\min}/2$. Every coordinate on the segment from $\widehat{p}$ to p_A is therefore at least $\widehat{p}_{\min}/2$. Coordinatewise Taylor expansion gives

$$\log(\widehat{p} + h) - \log \widehat{p} = \mathrm{diag}(\widehat{p})^{-1} h + r, \qquad \|r\|_2 \le \frac{2}{\widehat{p}_{\min}^2} \|h\|_2^2 .$$

Subtract the two KKT equations, write $h = Vz$, and rearrange:

$$z - J_{\widehat{A}}^{-1} V^\top H \widehat{p} = J_{\widehat{A}}^{-1} V^\top \{Hh - \tau r\}.$$

The max-norm bound implies $\|H\|_{\mathrm{op}} \le m\delta$. Combining it with $\|h\|_2 \le \delta/\tau$ proves

$$\left\| p_A - \widehat{p} - V J_{\widehat{A}}^{-1} V^\top H \widehat{p} \right\|_2 \le \widehat{\gamma} \frac{\delta^2}{\tau} \left(m + \frac{2}{\widehat{p}_{\min}^2} \right). \tag{B.20}$$

Consequently,

$$\|p_A - \widehat{p}\|_1 \le \min \left\{ 2, \frac{\delta}{\tau}, \Lambda(\widehat{A}, \widehat{p})\delta + \sqrt{m}\,\widehat{\gamma} \frac{\delta^2}{\tau} \left(m + \frac{2}{\widehat{p}_{\min}^2} \right) \right\}. \tag{B.21}$$

Outside this local radius, the first two global terms remain valid. □

Lemma B.5 (Estimated-game exploitability). *Let A and $\widehat{A}$ be skew-symmetric with $\left\|A - \widehat{A}\right\|_{\max} \le \delta$. Define the one-sided symmetric saddle gap*

$$\mathrm{Gap}_{\widehat{A},\tau}(\widehat{p}) = \max_{u \in \Delta_m} \left\{ u^\top \widehat{A} \widehat{p} - \tau\, \mathrm{KL}(u \| \pi_0) + \tau\, \mathrm{KL}(\widehat{p} \| \pi_0) \right\}.$$

If $\mathrm{Gap}_{\widehat{A},\tau}(\widehat{p}) \le \varepsilon_{\mathrm{opt}}$, then

$$\mathrm{Expl}_A(\widehat{p}) := \max_i e_i^\top A \widehat{p} \le \delta + \tau \log(1/\pi_{0,\min}) + \varepsilon_{\mathrm{opt}}.$$

Proof. By the definition of $\mathrm{Gap}_{\widehat{A},\tau}$, for every i,

$$\begin{aligned} e_i^\top \widehat{A} \widehat{p} &\le \tau \{ \mathrm{KL}(e_i \| \pi_0) - \mathrm{KL}(\widehat{p} \| \pi_0) \} + \varepsilon_{\mathrm{opt}} \\ &\le \tau \log(1/\pi_{0,\min}) + \varepsilon_{\mathrm{opt}}. \end{aligned}$$

Moreover,

$$e_i^\top A \widehat{p} \le e_i^\top \widehat{A} \widehat{p} + \left| e_i^\top (A - \widehat{A}) \widehat{p} \right| \le e_i^\top \widehat{A} \widehat{p} + \delta.$$

Maximizing over i proves the claim. □

Completion of Theorem 4.4. Let $\Delta = \widehat{\Theta} - \Theta^\star$. For $(q, r) \in S$, orthogonality of the Hodge spaces gives

$$\left\| \widehat{\theta}_{qr}^{\mathrm{hyb}} - \theta_{qr}^\star \right\|_2^2 = \|\Delta_{qr} P_{\mathsf{G}}\|_2^2 + \left\| \theta_{qr}^\star P_{\mathsf{C}} \right\|_2^2 ; \tag{B.22}$$

outside S the corresponding squared error is $\|\Delta_{qr}\|_2^2$. On simultaneous bracket coverage, every selected row obeys $\left\|\theta_{qr}^\star P_{\mathsf{C}}\right\|_2 = g_{qr}(\Theta^\star) \le U_{qr}$. Combining (B.22) with (4.10) therefore yields

$$\sum_r w_{qr} \left\| \widehat{\theta}_{qr}^{\mathrm{hyb}} - \theta_{qr}^\star \right\|_2^2 \le B_{q,w}^2 + \sum_{r \in \widehat{S}_q} w_{qr} U_{qr}^2 = \overline{R}_q^2.$$

If the component bounds in Theorem 4.4 are available, orthogonality also gives

$$\sum_r w_{qr} \left\| \widehat{\theta}^{\mathrm{hyb}}_{qr} - \theta^\star_{qr} \right\|_2^2 = \sum_r w_{qr} \left\| \Delta_{qr} P_{\mathsf{G}} \right\|_2^2 + \sum_{r \notin \widehat{S}_q} w_{qr} \left\| \Delta_{qr} P_{\mathsf{C}} \right\|_2^2 + \sum_{r \in \widehat{S}_q} w_{qr} g_{qr}^2$$
$$\leq B_{q,\mathsf{G}}^2 + \sum_{r \notin \widehat{S}_q} w_{qr} C_{qr}^2 + \sum_{r \in \widehat{S}_q} w_{qr} U_{qr}^2 = \mathfrak{R}_q^2.$$

Both bounds hold on the same event, proving the squared-error claim with $R_q^\sharp = \min\{\overline{R}_q, \mathfrak{R}_q\}$.

Put $d_{qre} = \widehat{\theta}^{\mathrm{hyb}}_{qre} - \theta^\star_{qre}$. Lipschitzness, Cauchy–Schwarz, and $\sum_r w_{qr} = 1$ give, for every edge e,

$$\left| \widehat{A}^{\mathrm{hyb}}_{q,e} - A^\star_{q,e} \right| \leq L_h \sum_r w_{qr} \left| d_{qre} \right| \leq L_h \left(\sum_r w_{qr} d_{qre}^2 \right)^{1/2} \leq L_h R_q^\sharp.$$

Lemma B.4 now proves the policy bound, and Lemma B.5 proves the exploitability bound, including the additive $\varepsilon_{\mathrm{opt}}$ term for an inexact saddle solution.

Let $\mathcal{E}_{\mathrm{aud}}$ be simultaneous rowwise bracket coverage and $\mathcal{E}_{\mathrm{rec}}$ the recovery event in (4.10), including the optional component bounds when the second radius term is used. On $\mathcal{E}_{\mathrm{aud}}$ every resolved label is correct. A union bound under arbitrary dependence gives

$$\mathbb{P}(\mathcal{E}_{\mathrm{aud}} \cap \mathcal{E}_{\mathrm{rec}}) \geq 1 - \mathbb{P}(\mathcal{E}_{\mathrm{aud}}^c) - \mathbb{P}(\mathcal{E}_{\mathrm{rec}}^c) \geq 1 - \delta - \eta,$$

simultaneously for every query. Finally, a global Frobenius event implies $\sum_r w_{qr} \left\| \Delta_{qr} \right\|_2^2 \leq (\max_r w_{qr}) \left\| \Delta \right\|_F^2$, proving the stated choice of $B_{q,w}$. □

Proposition B.6 (Three-response cycle)**.** *For $a, b, c > 0$, the strict-cycle game and its unique unregularized symmetric equilibrium are*

$$A = \begin{bmatrix} 0 & a & -c \\ -a & 0 & b \\ c & -b & 0 \end{bmatrix}, \qquad \pi^0 = \frac{(b, c, a)^\top}{a + b + c}. \tag{B.23}$$

Every pure policy has exploitability at least $\min(a, b, c)$, while $\mathrm{Expl}_A(\pi^0) = 0$. The regularized policy in (3.3) converges to π^0 as $\tau \downarrow 0$.

Proof of Proposition B.6. Direct multiplication gives $A(b, c, a)^\top = 0$, so π^0 guarantees value zero against every pure response; skew symmetry gives the converse guarantee. It is therefore a symmetric equilibrium. In a skew game the row- and column-optimal strategy sets coincide: row optimality is equivalent to $\max_i (Ap)_i \leq 0$. Since $p^\top Ap = 0$, every optimal strategy has $\max_i (Ap)_i = 0$. Every pure action is beaten by its cyclic predecessor, so equilibrium support has at least two actions. Each two-action support produces a positive payoff component: for $\{1, 2\}$, $(Ap)_1 = ap_2 > 0$; for $\{2, 3\}$, $(Ap)_2 = bp_3 > 0$; and for $\{1, 3\}$, $(Ap)_3 = cp_1 > 0$. Hence every equilibrium strategy has full support and complementary slackness gives $Ap = 0$. The null space of this nonzero 3×3 skew matrix is one-dimensional and is spanned by $(b, c, a)^\top$, proving uniqueness after normalization. For the pure policies e_1, e_2, e_3, exploitability is respectively c, a, b, proving the stated lower bound. Finally, let p_τ be the regularized symmetric policy. Row optimality against $s = p_\tau$ gives, for every pure e_i,

$$e_i^\top Ap_\tau \leq \tau \{ \mathrm{KL}(e_i \| \pi_0) - \mathrm{KL}(p_\tau \| \pi_0) \} \leq \tau \log(1/\pi_{0,\min}).$$

Every limit point therefore satisfies $\max_i e_i^\top Ap \leq 0$ and is an unregularized equilibrium. Uniqueness forces $p_\tau \to \pi^0$. □

C Experiment Details and Additional Audits

C.1 Theorem–Evidence Atlas

Figure 4 aligns each empirical object with its falsifiable claim: audit outcomes, score and coverage events, the exact oracle risk identity, selector disagreement, a real acquisition diagnosis, and nonlinear deployment sensitivity. The generating script exports vector PDF, a 300-dpi preview, a statistics report, and hashes of every input and output.

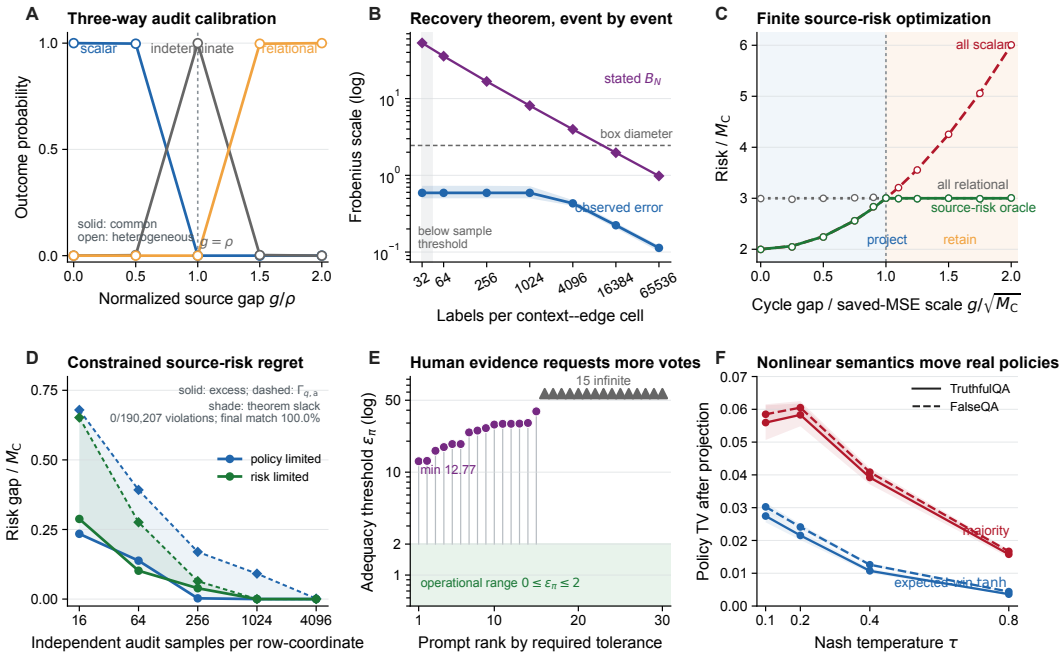


Figure 4: **Evidence follows the theory chain. A**, the three-way audit resolves both separated branches and retains the boundary under common and heterogeneous workers (Theorem 4.1 and Corollary 4.1). **B**, theorem-prescribed penalties satisfy the score and coverage events, and the stated envelope covers every realized recovery error (Theorem 4.2 and Corollary 4.2). **C**, Monte Carlo points coincide with the exact weighted source-logit risk and value of the finite hybrid optimization (Theorem 4.3). **D**, on 190,207 covered split-sample replications, conditional excess risk stays below the threshold-straddling remainder with zero violations under both policy- and risk-limited gates, and converges to the constrained source-risk oracle (Corollary 4.4). **E**, all 30 exact-cell human graphs require a policy tolerance above the operational range or have an infinite upper endpoint. **F**, two real sources show persistent policy movement under nonlinear majority and expected-win semantics across τ.

C.2 Direct finite-sample test check

The script `experiments/scalarity_test_calibration.py` evaluates both branches on the balanced three-edge graph with $a = .2$, $\delta = .05$, and 20,000 independent replications per condition. For $n \in \{128, 512, 1024, 2048, 4096\}$, the analytic test has zero null rejections and powers $.0346, 1, 1, 1, 1$ against cycle norm 2.2; its sufficient power condition first holds at $n = 1024$. The exact confidence-box test has null rejection between zero and 5×10^{-5} and power one throughout. The simulation therefore records both the guaranteed sufficient boundary and the practical gain from exact binomial inversion.

A positive-tolerance block fixes $\rho = .6$, adds a nonzero gradient potential, uses unbalanced cell sizes $(8192, 16384, 32768)$, and crosses residual ratios $0, .5, 1, 1.5, 2$ with common-probability and independent four-stratum heterogeneous workers. Each of the ten cells has 20,000 replications. Bracket coverage is one and the wrong-resolved rate is zero in every cell. For common probabilities, the respective $(\text{adequate}, \text{indeterminate}, \text{nonintegrable})$ rates are $(1, 0, 0)$, $(.99785, .00215, 0)$, $(0, 1, 0)$, $(0, .00275, .99725)$, and $(0, 0, 1)$; the heterogeneous rates differ by at most .0005. Thus the experiment exercises both resolved branches, the boundary, unequal information, and the sampling model used by Theorem 4.1.

C.3 Audit-driven multi-rubric hybrid

The script `experiments/hybrid_audit_sim.py` fixes six queries, five rubrics, and a complete three-response graph. Its all-adequate, mixed, and boundary profiles use cycle-to-margin ratios $(0, .1, .15, .2, .25)$, $(.1, .25, 1.55, 1.9, 2.3)$, and $(.1, .25, 1, 1.55, 2.1)$ at $\rho = .7$. Every profile

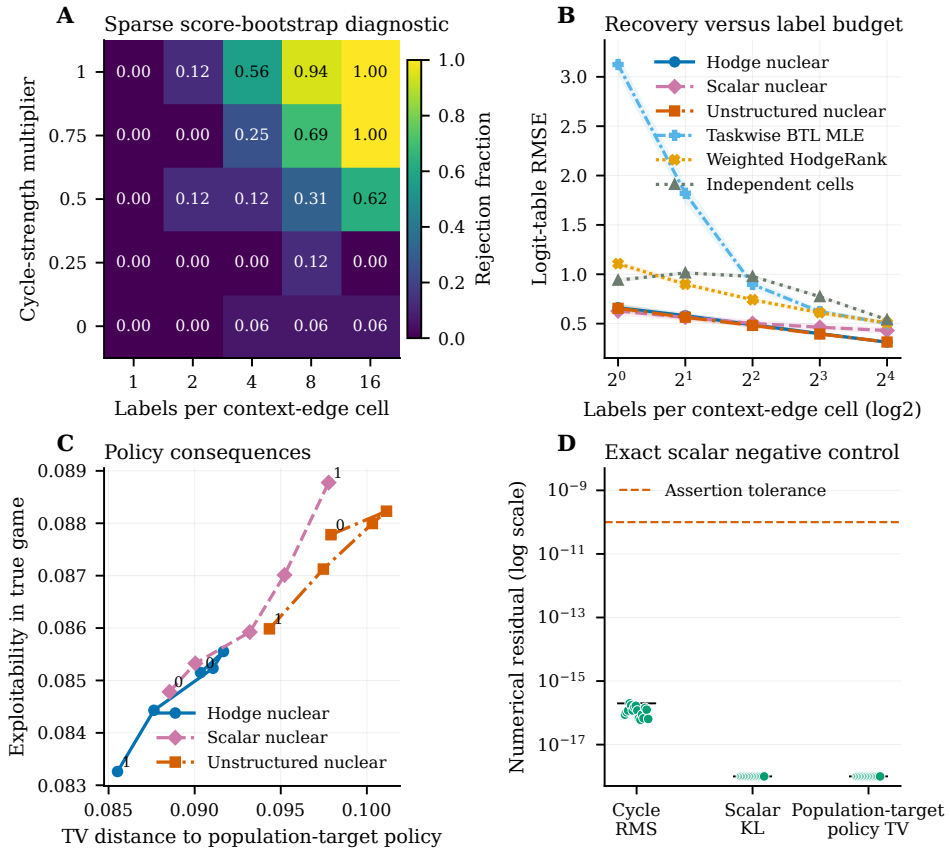


Figure 5: Full controlled-geometry recovery study. **A**, a score bootstrap diagnoses the strongest cycle. **B**, recovery traces the efficiency gain of scalar restriction at $c = 0$ and its growing bias as cycle strength increases. **C**, policy total variation and true-game exploitability quantify decision consequences. **D**, the exact scalar population control verifies zero structural residual. Ribbons are 95% Student-t Monte Carlo intervals over 16 latent problems.

therefore contains positive-margin rows with $0 < g < \rho$. Each of 300 replications forms one Bonferroni–Clopper–Pearson rectangle over all 90 query–rubric–edge cells and inverts the triangle cycle coordinate exactly. Every query deploys: certified adequate rows project, while nonintegrable and indeterminate rows retain their relational estimates. Resolution remains a diagnostic, and a deployment may replace the relational fallback with acquisition or abstention. Identity deployment isolates adequacy-based representation routing; the UltraFeedback experiment evaluates nonlinear deployment semantics.

All 5,400 query–replication instances deploy; unresolved rows use the relational fallback. Performance first averages over all six queries within each replication and then gives equal weight to replications. At $n = 4096$, the all-adequate and mixed profiles resolve every query, while all 600 boundary-profile instances deploy through the fallback, giving overall resolution .667. The mixed profile projects two rows and retains three; the boundary profile projects two rows and retains the two separated relational rows plus the row exactly at the margin. The plug-in and forced exact-equality rules disagree with the population-adequacy selector on 16.7% and 57.9% of queries, respectively. The audit projects 60% of rows; its mean weighted source error is .1050 against mean $R_q^\sharp = .3041$, with payoff error .0486 and policy TV .0147. Mean weighted projection bias .09194 induces .03630 max-payoff bias and .01041 policy TV. Its payoff error is 85.2% below always scalar (.3288). At each budget, the family-wise wrong-row, wrong-query, joint source-certificate, joint payoff-certificate, and joint policy-certificate failure counts are all $0/300$ over the complete six-query family, with 95% Clopper–Pearson upper endpoint .0122. The manifest records the fixed population, seeded design, package versions, script digest, and output hashes.

Panel A: all-query performance with relational fallback

n	Method	N_q	Audit resolution	Source error	$R_q^\sharp$	Payoff error	Policy TV	TV radius
1024	Audit hybrid	1800	0.486	0.1363	0.551	0.0608	0.0233	1.000
1024	Plug-in hybrid	1800	0.486	0.1711	–	0.0734	0.0247	–
1024	Exact-equality test	1800	0.486	0.1364	–	0.0610	0.0233	–
1024	Population adequacy	1800	0.486	0.2033	–	0.0866	0.0259	–
1024	Always scalar	1800	0.486	0.6993	–	0.3413	0.0893	–
1024	Always relational	1800	0.486	0.1101	–	0.0388	0.0201	–
2048	Audit hybrid	1800	0.663	0.1171	0.405	0.0541	0.0182	0.960
2048	Plug-in hybrid	1800	0.663	0.1507	–	0.0655	0.0195	–
2048	Exact-equality test	1800	0.663	0.1168	–	0.0536	0.0182	–
2048	Population adequacy	1800	0.663	0.1886	–	0.0798	0.0211	–
2048	Always scalar	1800	0.663	0.6930	–	0.3344	0.0870	–
2048	Always relational	1800	0.663	0.0782	–	0.0280	0.0145	–
4096	Audit hybrid	1800	0.667	0.1050	0.304	0.0486	0.0147	0.760
4096	Plug-in hybrid	1800	0.667	0.1424	–	0.0610	0.0164	–
4096	Exact-equality test	1800	0.667	0.0938	–	0.0406	0.0134	–
4096	Population adequacy	1800	0.667	0.1799	–	0.0743	0.0182	–
4096	Always scalar	1800	0.667	0.6896	–	0.3288	0.0852	–
4096	Always relational	1800	0.667	0.0547	–	0.0190	0.0099	–

Panel B: replication-level simultaneous checks

n	Reps	N_{res}	Wrong row	Wrong query	Source fail	Payoff fail	TV fail	Box coverage
1024	300	875	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	291/300 [0.944, 0.986]
2048	300	1194	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	285/300 [0.919, 0.972]
4096	300	1200	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	0/300 [0.012]	290/300 [0.940, 0.984]

Panel A first averages over all six registered queries within each replication and then gives equal weight to replications; N_q counts query–replication instances. Audit resolution is diagnostic: unresolved rows retain their relational estimate. $R_q^\sharp$ is the source and payoff radius for the audit hybrid, and the displayed TV radius is $\min\{1, R_q^\sharp/(2\tau)\}$. N_{res} counts fully resolved query–replication instances. Panel B reports family-wise counts over the complete six-query family in every replication. For zero counts, brackets give the upper endpoint of a two-sided 95% Clopper–Pearson interval; box coverage shows the full interval.

Table 3: Audit-driven hybrid performance under the conservative relational fallback in Theorem 4.4. Panel A compares all methods on the same complete query family; payoff error is max norm and policy error is total variation. $R_q^\sharp$ certifies source and payoff error for every audit-hybrid query, including unresolved queries. Panel B reports replication-level family-wise errors, all-query certificate failures, and simultaneous-box coverage with exact 95% intervals.

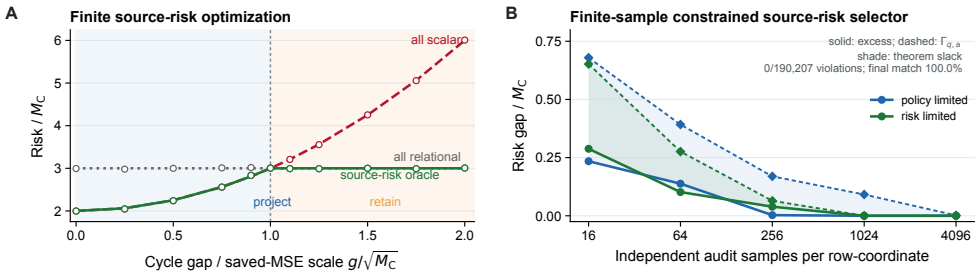


Figure 6: Finite source-risk optimization and split-sample calibration. Panel A verifies the R threshold comparisons and their strict mixed-regime gain over both uniform representations (Theorem 4.3); panel B reports conditional excess risk and the threshold-straddling remainder under policy- and risk-limited gates (Corollary 4.4).

C.4 Split-sample audit-to-oracle calibration

The script `experiments/audit_oracle_calibration.py` isolates Corollary 4.4 in its known-M_C Gaussian sequence model. It crosses binding policy and risk gates with $n \in \{16, 64, 256, 1024, 4096\}$ and 20,000 replications per cell. Independent audit draws construct simultaneous Bonferroni Gaussian brackets; independent recovery draws evaluate the selected representation. All 190,207 bracket-covered replications contain zero unsafe projections and zero violations of (4.9). Conditional excess risk is zero from $n = 1024$ onward; at $n = 4096$, the largest covered mean remainder is $1.08 \times 10^{-4} < .002 M_C$. Direct recovery-block risk estimates agree with the conditional risk calculation within 1.34 Monte Carlo standard errors. This experiment targets the risk selector with known component MSE; the fixed-ρ Bernoulli study targets adequacy and abstention, and the matrix experiments target recovery.

Method	Weighted source risk	Payoff ℓ_∞	Policy TV	Unsafe (%)	Oracle match (%)	Resolved (%)
Risk-gated hybrid	0.612	0.131	0.083	0.0	100.0	76.2
Always scalar	1.001	0.218	0.103	–	–	–
Always relational	0.799	0.148	0.084	–	–	–
Adequacy only	0.692	0.150	0.087	24.8	75.2	100.0
Plug-in selector	0.790	0.148	0.084	7.6	72.8	62.5
Generic low-rank	0.906	0.154	0.086	–	–	–

Table 4: End-to-end Bernoulli comparison. Unsafe projection, registry-oracle agreement, and resolution apply to selectors; lower source risk, payoff error, and policy TV are better. Values are paired-replication means; intervals and Holm-adjusted tests are archived separately.

C.5 End-to-end Bernoulli selector stress test

The script `experiments/end_to_end_hybrid.py` joins the full chain in one paired Bernoulli experiment. Every query contains scalar, weak-cycle, medium-cycle, and strong-cycle rubrics. An audit block constructs a Bonferroni–Clopper–Pearson rectangle over all query–rubric–edge cells and applies the finite rectangular inversion in Proposition B.2; an independent recovery block estimates the gradient and cycle components; a fixed, independently generated cyclic-MSE registry supplies the risk threshold. The Hodge rank truncation is a fixed estimator for testing the exact source-risk identity and selector in Theorem 4.3 and Corollary 4.4. The convex estimator in (4.2) is evaluated separately in Figure 5 and the explicit-bound calibration below.

All methods share the same audit and recovery draws. The risk-gated hybrid is compared with always-scalar and always-relational representations, an adequacy-only ablation that removes the MSE gate, a plug-in selector that removes honest confidence regions, and an undecomposed rank-four recovery. Source risk, payoff max error, policy total variation, regularized saddle gap, unsafe projection, registry-oracle agreement, and certified resolution are computed on every paired replication. Sensitivities vary the audit–recovery allocation, policy tolerance, audit budget, and cyclic-MSE multiplier.

Across 96 independent vote realizations of one fixed six-query population, the estimated hybrid weighted source risk is .6119 (95% CI .5898–.6339), a 38.9% reduction from always scalar and a 23.4% reduction from always relational. Each replication records $Q^{-1}\sum_q\sum_r w_{qr}\|\widehat{\theta}_{qr}-\theta^\star_{qr}\|_2^2$; the replication mean estimates the query-averaged risk in Theorem 4.3. Its separate unweighted per-cell MSE is .1012. Every covered gate avoids unsafe projection; the deployed action, including relational defaults for unresolved rows, agrees with the registry-defined constrained oracle on 100% of rows, and 76.2% of rows resolve. The hybrid has lower payoff max error than every baseline (largest Holm-adjusted one-sided $p = 3.03 \times 10^{-21}$) and lower weighted source risk than every baseline (largest adjusted $p = 6.93 \times 10^{-35}$). Its policy TV is .0828: the paired improvement is significant against always scalar, adequacy only, and generic low rank, while the always-relational and plug-in contrasts have adjusted $p = .0986$ and .0785. The exact audit rectangle covers globally in all 96 replications; this observed rate is a calibration diagnostic for the prespecified familywise noncoverage level .10. The estimated-M_{C} sensitivity attains 99.3% interval coverage of the independent registry target.

C.6 Complementary Hodge rank-mismatch phase diagram

The phase experiment fixes $Z = 36$, $m = 5$, total component rank four, logit bound 1.8, estimator bound 6, and $\tau = .16$. It crosses $(r_{\mathsf{G}}, r_{\mathsf{C}}) \in \{(3,1),(2,2),(1,3)\}$, cycle scale in $\{.35, 1\}$, and expected labels per cell in $\{.25, .5, 1, 2, 4, 8, 16\}$. Twenty-four independent latent problems per structural condition use paired nested missing-at-random outcomes. Penalties are calibrated by the same 32-draw design-conditional Rademacher procedure for both estimators. All 2,016 fits converge; the largest proximal-gradient residual is 1.80×10^{-7}.

Generic low rank has lower total logit RMSE in all 42 cells by .43–4.64%. Hodge has lower population-target policy TV in 36 cells, in all 30 cells with at least one expected label per cell, and in all 18 cells from budget four onward with cellwise paired 95% intervals excluding zero. These intervals are descriptive and unadjusted for multiplicity. The manifest records every paired draw, convergence statistic, package version, and output hash.

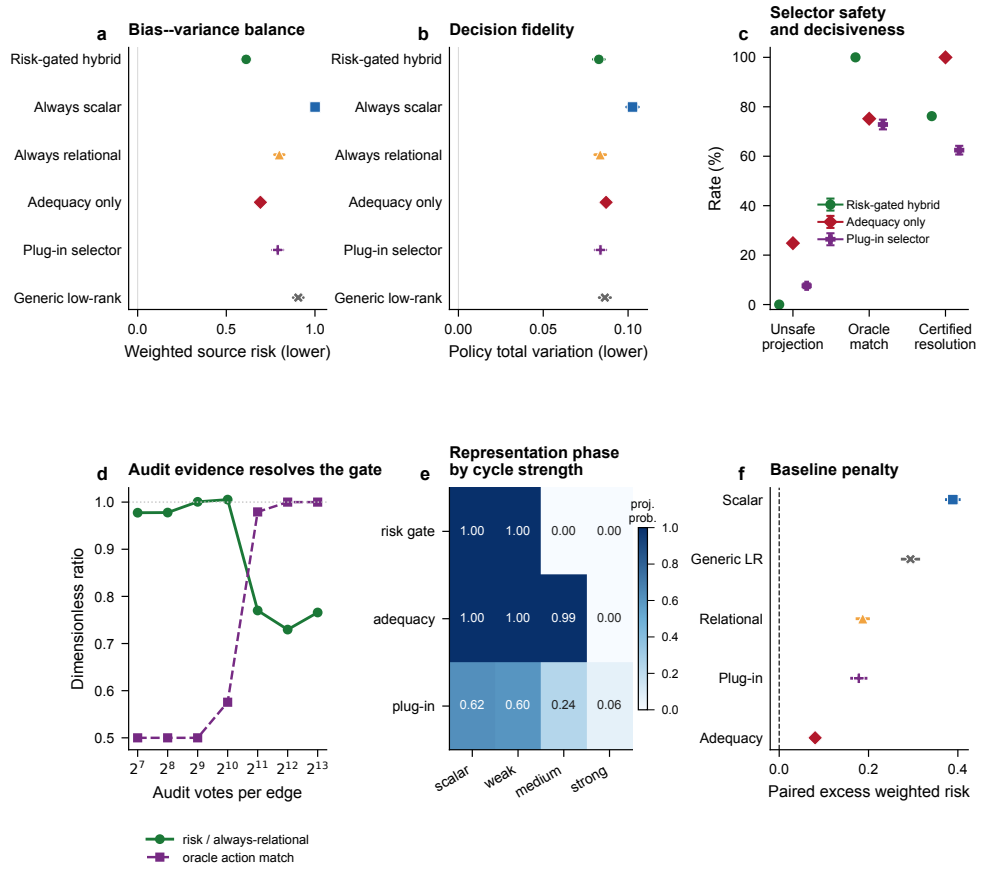


Figure 7: Panels a and f test the projection-bias versus cyclic-MSE comparison in Theorem 4.3; panels c–e test confidence-based selection, the indeterminate branch, and convergence to the registry-defined constrained oracle in Corollary 4.4; panel b propagates the selected table to the policy target in Theorem 4.4. Points and intervals use paired replications.

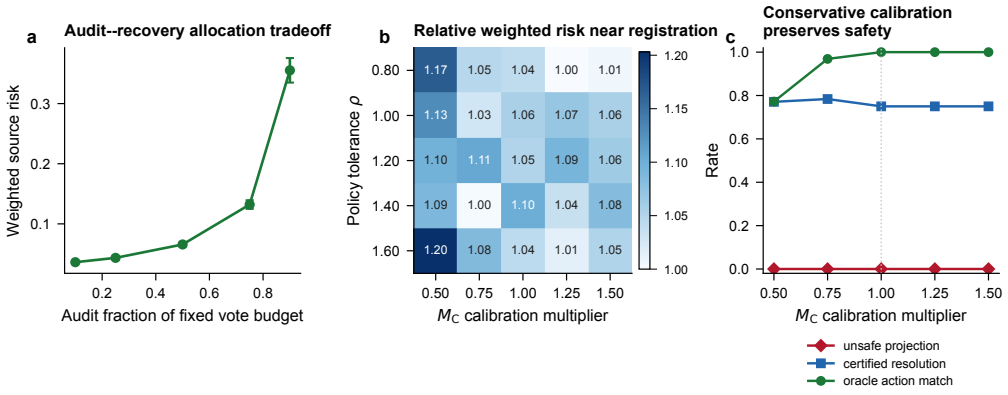
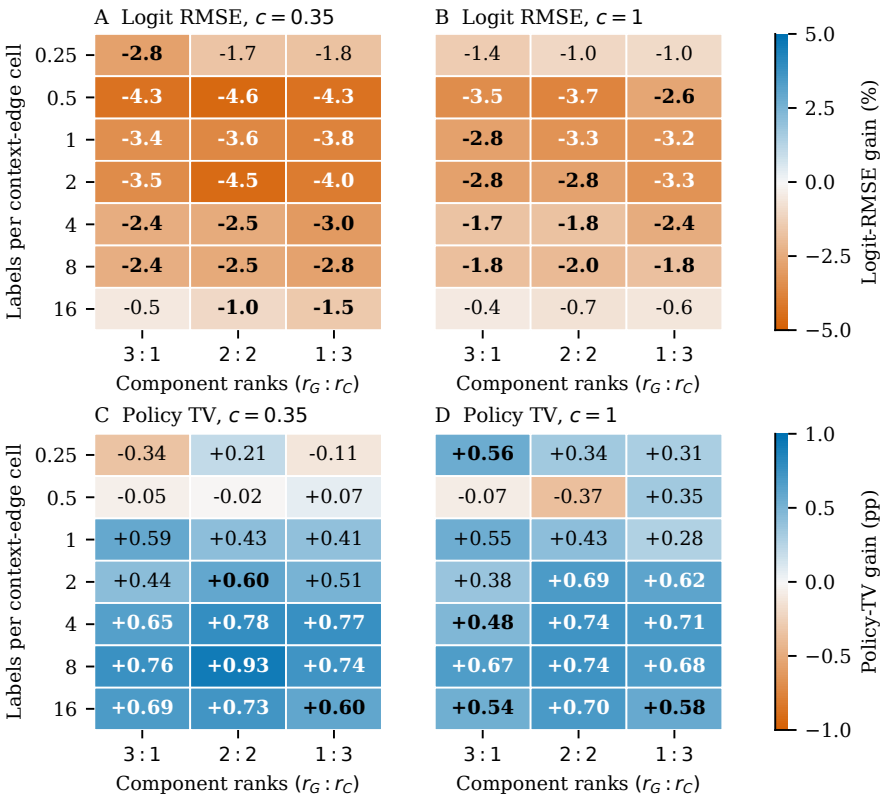


Figure 8: Registered sensitivity analysis. Panel a reallocates a fixed vote budget between audit and recovery, panel b crosses policy tolerance with the cyclic-MSE calibration multiplier, and panel c displays the safety–resolution tradeoff. The main certificate uses the fixed independent MSE registry; estimated-MSE perturbations are reported as robustness diagnostics.



Positive (blue) favors Hodge; negative (orange) favors generic low rank.
Bold: cellwise, unadjusted paired 95% Student-t interval excludes zero.

Figure 9: The Hodge split reallocates error toward decision-relevant directions. **A–B**, generic nuclear recovery has lower total logit RMSE in 42/42 cells. **C–D**, Hodge recovery has lower population-target policy TV in 36/42 cells and all 30 cells with at least one label per context–edge cell. Total rank is fixed at four; each cell averages 24 paired latent problems. Positive values favor Hodge, and bold values mark unadjusted paired 95% Student-t intervals excluding zero.

C.7 Theory-matched Bernoulli simulation

The simulation uses Z context rows and the complete comparison graph on m responses. Its construction avoids basis-dependent random-number consumption from an arbitrary SVD basis of a repeated singular subspace. It generates raw low-rank matrices $U_G V_G^\top$ and $U_C V_C^\top$, right-projects them with P_G and P_C, and applies a single global scale factor to enforce $\|\Theta\|_{\max} \le \alpha$. The script asserts the requested component ranks, orthogonality, and machine-zero cycle energy at the scalar negative control before drawing any outcomes.

Truth is globally rescaled to $\alpha_0 = 1.8$; every nuclear estimator uses the wider box $\alpha = 6$. For each observed (x, e) cell, votes are Bernoulli with probability $\sigma(\theta_{xe})$. The solver evaluates (4.2), computes the full constrained proximal map by Dykstra's algorithm as in Section B.3, and backtracks until the composite descent inequality holds. Stopping requires both relative iterate change and a normalized first-order residual below the recorded tolerances.

The comparison methods share the same observations and optimization budget:

1. **Hodge low rank**: estimator (4.2);
2. **scalar low rank**: the same solver constrained to $\Theta P_C = 0$;
3. **unstructured low rank**: one nuclear penalty on Θ;

Loss	PCWST−Hodge [95% CI]	Cond. CI wins	Paired wins	Holm p
Logit RMSE	0.1725 [0.1462, 0.1988]	9/9	141/144	$< 10^{-4}$
Population KL	0.0201 [0.0161, 0.0241]	9/9	141/144	$< 10^{-4}$
Brier loss	0.0083 [0.0069, 0.0097]	9/9	141/144	$< 10^{-4}$
Payoff RMSE	0.0186 [0.0139, 0.0233]	9/9	132/144	$< 10^{-4}$
Policy TV	0.0218 [0.0146, 0.0289]	9/9	128/144	$< 10^{-4}$
Exploitability	0.0010 [-0.0016, 0.0037]	0/9	73/144	0.2086

Table 5: Paired comparison using the pinned official PCWST solver with a registered shared-scale adapter. Positive PCWST-minus-Hodge differences favor Hodge because every row is a loss. Condition intervals are paired 95% Student-t intervals; omnibus tests first average the nine repeated conditions within each of 16 independent latent problems, then apply Holm correction across the reported losses.

4. **taskwise BTL MLE**: independent bounded likelihood fits in the gradient space;
5. **weighted HodgeRank**: rowwise weighted least-squares projection of Jeffreys-smoothed edge logits;
6. **independent cells**: clipped empirical cell logits with no pooling.

For each replicate and observation budget, the three penalties are calibrated as 0.5 times the empirical 90th percentile of the corresponding projected score norm over 32 design-conditional Rademacher draws. This outcome-blind calibration depends only on the sampled design. Metrics are evaluated against the registered population logits, payoffs, and policies. The archived manifest records the calibration settings, seeds, projected ranks, solver diagnostics, package versions, and wall time.

At 16 labels per cell and zero cycle strength, scalar and Hodge nuclear recovery have RMSE .206 and .244; taskwise BTL and weighted HodgeRank have RMSE .338 and .331. The scalar/Hodge ordering reverses by cycle multiplier .5 on the tested grid. At multiplier one, Hodge RMSE is .312 versus .430 for scalar nuclear, .504 for taskwise BTL, and .507 for weighted HodgeRank; policy-TV errors for Hodge/scalar are .086/.098. The strongest-cycle score bootstrap resolves 16/16 replicates, and the scalar population negative control has maximum structural residual 1.99×10^{-16}. These paired results identify the efficiency-to-bias transition and separate cross-row low-rank sharing from rowwise scalar estimation.

The official PCWST solver at commit `ee56a27`, coupled to a registered shared-scale adapter, adds a recent general skew-matrix comparator (Lee & Chen, 2026). Each prompt–rubric row is converted to its response-by-response win-count matrix; one scale for the full simulated table is selected by aggregate 20% held-out Bernoulli likelihood, and every row is refit on all votes. Across 16 paired latent problems, three cycle strengths, and budgets four, eight, and sixteen, all 144 replicate–condition tables have fully converged selected-scale tuning fits and all-vote refits (5,184 rowwise fits at each stage; maximum 80 and 112 iterations, respectively, below the 1,000-iteration cap). Hodge recovery has lower mean logit RMSE, KL, Brier loss, payoff RMSE, and policy TV in all nine conditions. Averaging the repeated grid within each latent problem gives Holm-adjusted one-sided $p < 1.7 \times 10^{-5}$ for all five losses. The registered exploitability contrast is inconclusive after Holm correction ($p = .209$), whereas the five estimation and policy-fidelity losses separate.

Under the upstream-aware convergence criterion (exit code zero and termination strictly before the pinned 1,000-iteration cap), the official per-row tuning routine returns finite full tables in 113 of 144 cells and fully converged tables in 6 of 144 cells, spanning three latent problems; no solve hits the cap, and the maximum tuning and refit counts are 628 and 463. Replicate-aggregated point estimates on the finite-return subset favor Hodge, conditional on solver return; the fully converged subset is reported descriptively. The shared-scale adapter supplies the complete 144-cell paired comparison in Table 5. Scripts, convergence summaries, and both sensitivity subsets are archived with the main comparison.

The two sparsest grid points occupy the tolerance regime of Corollary 4.3. With $Z = 36$, $E = 10$, and $D = 360$, the three Hodge tangent dimensions are 152, 156, and 156. At budgets .25 and .5

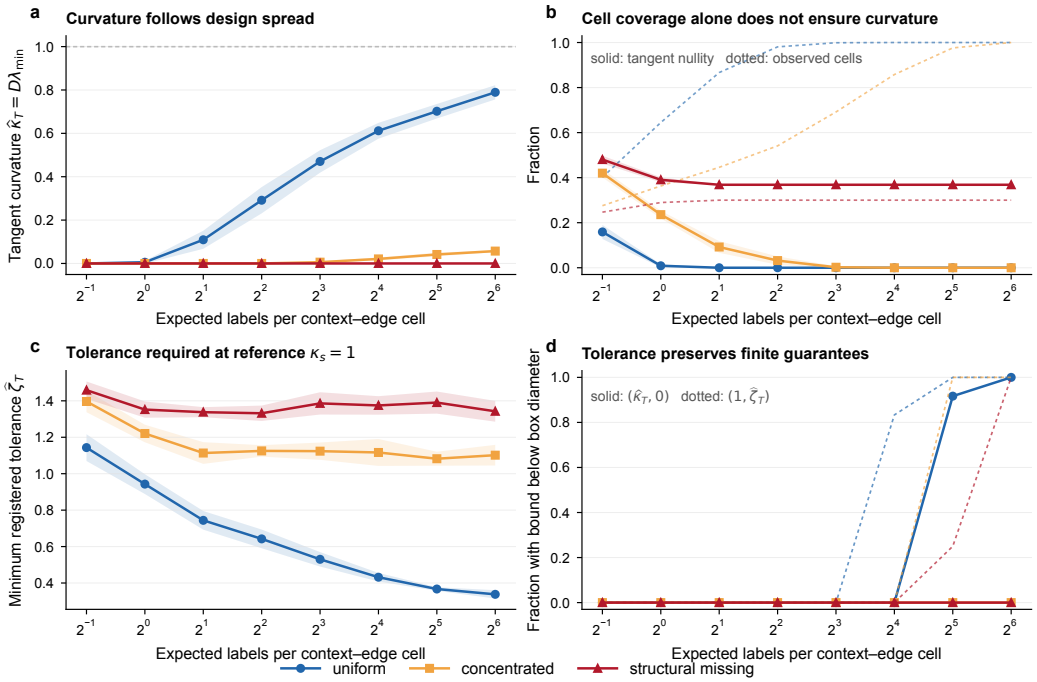


Figure 10: **Design geometry determines sparse recovery. a**, normalized minimum curvature on the registered Hodge tangent space. **b**, tangent nullity (solid) and observed-cell fraction (dotted) separate coverage from curvature. **c**, the tolerance required by the registered bounded tangent family at reference $\kappa_s = 1$. **d**, the fraction of replications in which the tangent-curvature envelope (solid) or tolerance-family envelope (dotted) is tighter than the box-diameter bound. Curves are means over 12 paired latent problems; ribbons are 95% Student-t intervals. The diagnostic targets the sampling mechanism in Corollary 4.3; its full weighted-cone condition is (B.11).

labels per cell, 76–85 and 134–149 distinct cells are observed. Each sampling operator has a nonzero tangent-space kernel, which precludes uniform zero-tolerance Frobenius RSC. The policy changes at these budgets are finite-dimensional measurements. A full-matrix guarantee begins when the tolerance becomes informative. The formal curvature uses the fitted box $\alpha = 6$, giving $q_6 = .00247$.

C.8 Sparse-design curvature and tolerance diagnostic

The script `experiments/sparse_rsc_diagnostic.py` isolates the design premise of Corollary 4.3. Twelve paired latent problems cross uniform sampling, 90% concentrated sampling, and permanent 30% structural support over eight budgets from .5 to 64 expected labels per cell. For each realized operator it computes the exact minimum eigenvalue and nullity on the truth's Hodge low-rank tangent space. It also computes the smallest tolerance over a registered bounded family containing every empirical tangent eigenvector and 256 seeded isotropic tangent directions. This finite-family diagnostic measures the mechanism in (B.11); the theorem's full weighted error-cone condition remains the formal assumption.

Uniform sampling eliminates tangent nullity by two labels per cell and raises normalized curvature from zero to .789 at budget 64. Concentrated sampling eventually covers every cell but reaches only .057 curvature, showing that raw coverage cannot replace spread. Structural missingness retains zero curvature and at least 36.8% tangent nullity at every budget. Whenever either registered numerical envelope becomes informative relative to the deterministic box-diameter bound, it covers the realized estimation error in 100% of checked cases (23 tangent-curvature checks and 73 tolerance-family checks). These results distinguish three regimes: positive curvature, delayed curvature with a finite tolerance, and permanent nonidentification requiring target restriction, acquisition, or abstention.

C.9 Explicit oracle-bound calibration

The script `experiments/oracle_bound_calibration.py` directly instantiates Theorem 4.2 and Corollary 4.2 with $Z = 2$, $E = 3$, component ranks $(1, 1)$, estimator box $\alpha = .5$, uniform random design, and the prescribed penalties $\lambda_j = 2u_j$. Across 128 latent problems and seven budgets from 32 to 65,536 labels per cell, it records both projected score inequalities, the multinomial minimum-count event, proximal optimality, realized Frobenius error, and B_N. The corollary's sample-size threshold is 43.85 labels per cell; every one of the 6×128 above-threshold runs satisfies the joint score-and-coverage event, and every realized error lies below B_N. The maximum error-to-bound ratio is .133. Over the final three budgets, the empirical error and stated envelope have log–log slopes $-.488$ and $-.505$, respectively. The envelope first beats the box-diameter bound at 16,384 labels per cell, quantifying both the theorem's coverage and its finite-sample conservatism.

C.10 Exact repeated-cell human comparison audit

The theorem-matched real-data experiment uses the human comparisons released with OpenAI Summarize-from-Feedback (Stiennon et al., 2020; OpenAI, 2020). We fix the `batch10/valid1` stratum before reading outcomes. The source is the validation Parquet at conversion commit `a4d82a1068e43c8edee69c452a7e763722bd5496`; its size is 22,831,890 bytes and its SHA-256 digest is

`e38618270bbc24d35da7b03671a6a701d20efbd4c96a6a853e0096ab85b8e4de`.

A comparison cell is keyed by batch, split, official source identifier, exact source text, and the unordered pair of exact stripped candidate texts. We collapse concordant duplicate records from one worker within a cell and remove a worker–cell group when its duplicate choices conflict. The outcome-blind graph rule retains an edge when it has at least ten distinct workers and participates in a complete candidate triangle. Applied to 7,084 source rows in the fixed stratum, the rule yields 118 complete triangles, 183 unique cells, 3,219 votes, 30 prompt graphs, and 26 workers. Retained cells contain 10–21 distinct workers.

For the common-probability analysis, each cell receives a two-sided Clopper–Pearson interval with Bonferroni error allocation over all 183 cells. We map probability endpoints through the extended logit and solve a linear feasibility problem against the gradient space separately for each prompt graph. Every confidence box intersects its gradient space, so the simultaneous familywise decision at $\alpha = .05$ and $\rho = 0$ is indeterminate. We also invert a prompt-local Bonferroni box, obtain a valid conservative finite-sample p-value by inversion, and apply Holm correction across the 30 graphs; this procedure rejects zero graphs, with minimum adjusted p-value one.

The heterogeneous-worker analysis treats the retained vote in cell d from worker k as an independent $\mathrm{Bernoulli}(p_{dk})$ variable and targets $\bar{p}_d = n_d^{-1} \sum_k p_{dk}$. Simultaneous two-sided Hoeffding intervals use radius $\sqrt{\log(2D/.05)/(2n_d)}$ over $D = 183$ cells. Their radii range from .4603 to .6670, and every one of the 30 prompt boxes intersects the gradient space. Thus worker heterogeneity does not change the resolved state. Worker identifiers enforce one retained vote per worker–cell. The same worker can label multiple cells; marginal interval coverage and the union bound permit the resulting cross-cell dependence.

For policy-scale resolution, we set $\tau = .2$ and predeclare $\rho = 2\tau\varepsilon_\pi = .4\varepsilon_\pi$. The grid contains 81 values over the full simplex ℓ_1 range $\varepsilon_\pi \in [0, 2]$. At every grid point the exact global boxes yield 0 scalar-adequate, 30 indeterminate, and 0 nonintegrable prompt graphs. The safe upper endpoint is finite for 15 graphs and infinite for 15 graphs because unanimous cells create infinite Clopper–Pearson logit endpoints. Among the finite cases, the first adequacy threshold is $\varepsilon_\pi = 12.7657$, outside the operational range. The curve is therefore a quantitative resolution diagnosis: it records the tolerance path and leaves both resolved branches uncertified at the current annotation budget.

The 10,000-draw parametric bootstrap is a complementary calibration diagnostic. It samples conditionally independent cells under prompt-specific Bradley–Terry fits and compares aggregate likelihood-ratio deviance. The observed deviance is 59.01 on 90 cycle degrees of freedom, at bootstrap lower-tail fraction .00675 and upper-tail $p = .9933$. For descriptive plug-in summaries, with k_e wins in the recorded orientation among n_e retained votes, we set $\widehat{\theta}_e = \log\{(k_e + 1/2)/(n_e - k_e + 1/2)\}$, project each retained triangle's edge-logit vector into gradient and cycle components, and map the

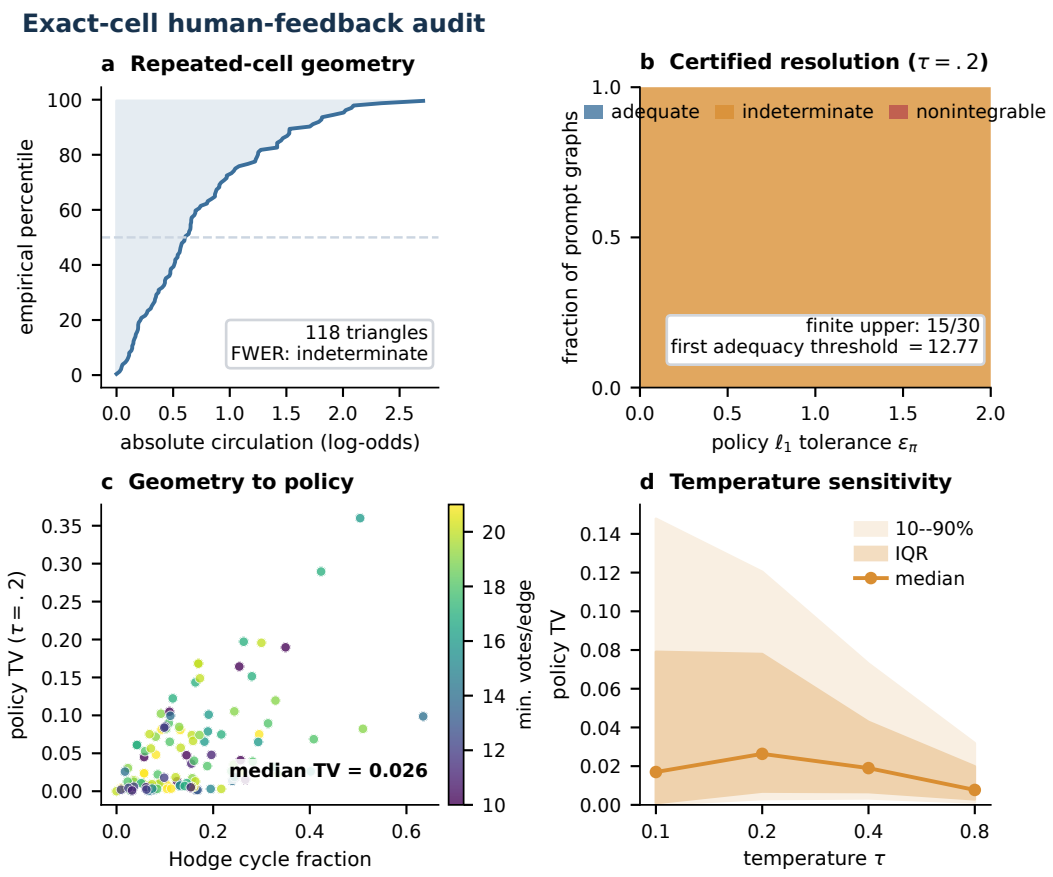


Figure 11: Full exact-cell audit. Panel a summarizes overlapping-triangle geometry. Panel b gives the simultaneous exact three-way decision over the policy-tolerance path. Panels c–d quantify plug-in decision sensitivity and temperature dependence. The complementary independent-cell bootstrap remains a secondary calibration diagnostic reported in text and machine-readable output.

original and gradient-projected logit tables through $\tanh(\cdot/2)$ before solving the three-action policies. Across the 118 overlapping triangles, the median absolute circulation is .605 log-odds and the median cycle fraction is .131. At $\tau = .2$, the policy total-variation shift has median .026, 90th percentile .120, and maximum .360. The run manifest records the selected-cell counts, prompt decisions, bootstrap summary, package versions, script digest, and hashes of the reported output files; the hashes bind these artifacts to the analyzed run.

C.11 Exploratory Arena aggregate audit

We additionally analyze the public Arena Human Preference 55K release (Chiang et al., 2024; LMArena, 2024). A fixed graph rule selects the largest complete model core with at least 100 decisive votes per edge: six models, 15 edges, and 5,123 battles. Each edge pools battles involving different prompts and generated responses, and the release provides no worker identifier. The resulting target is an aggregate model-pair relation under a stable i.i.d. routed-prompt-mixture and independent-battle working model. Its simultaneous confidence box intersects the scalar space; a fitted parametric bootstrap gives deviance 26.18 and $p = .0030$. The aggregate Hodge fraction is 25.6%, and projecting it away changes the policy by total variation .046 at $\tau = .2$. This aggregate analysis measures model-level mixture sensitivity; the exact-cell Summarize-from-Feedback study matches the repeated-Bernoulli audit assumptions.

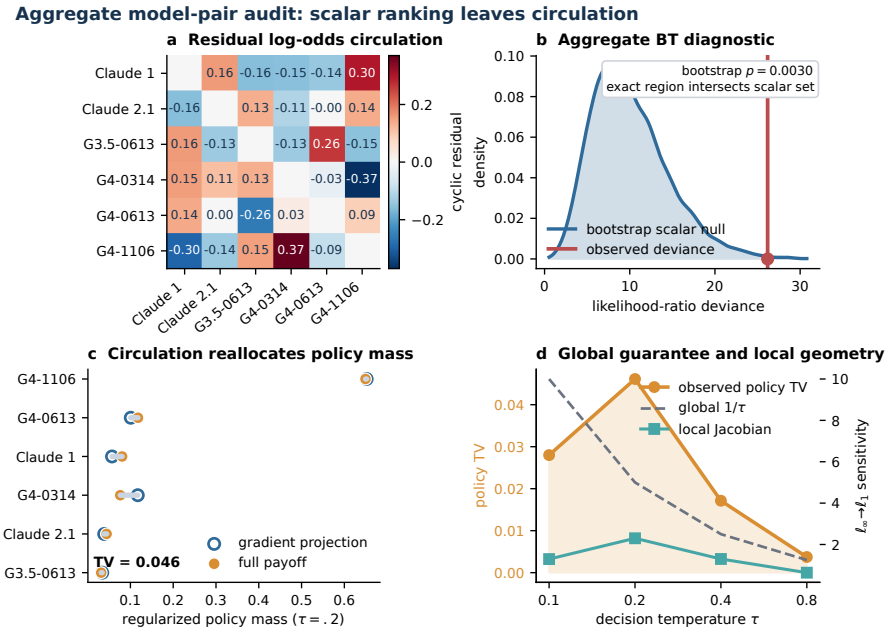


Figure 12: Exploratory Arena model-pair audit. Edges aggregate across prompts and response texts. The plot reports routed-mixture geometry, bootstrap calibration, policy reallocation, and local sensitivity under that aggregate target.

Source	Aggregation	n	Nonadditive	Strict cycle	Policy TV
TruthfulQA	Rubric majority	798	82.8%	0.00%	0.058
TruthfulQA	Nonlinear tanh	798	84.2%	0.00%	0.022
FalseQA	Rubric majority	2,317	87.1%	0.04%	0.061
FalseQA	Nonlinear tanh	2,317	88.7%	0.09%	0.024

Table 6: Observed downstream-payoff nonadditivity and policy consequences. "Nonadditive" denotes positive Hodge cycle energy above 10^{-12} after deterministic score aggregation. This table concerns aggregation nonadditivity; the separate Summarize-from-Feedback experiment supplies the repeated-Bernoulli scalarity test.

C.12 UltraFeedback processing and exclusions

We use the public `openbmb/UltraFeedback` dataset, which contains 63,967 prompts and four responses per prompt (Cui et al., 2024; OpenBMB, 2023). The audit covers all 811 TruthfulQA-source and 2,339 FalseQA-source rows. Thirteen TruthfulQA and twenty-two FalseQA records lack a complete numeric 4×4 rating array and are excluded before any statistic is computed, leaving 798 and 2,317 records.

The source JSONL files have SHA-256 digests

Slice	SHA-256	Rows
TruthfulQA	`e3f3281de0cfe906...56fff3de`	811
FalseQA	`b6ad211ac4ddef22...5180901`	2,339

The fetch script and run manifests record and verify every digest and source URL. The archive stores derived artifacts; reproduction downloads the public source slices on demand.

Table 6 records the exact source-wise estimates behind Figure 13.

For prevalence intervals, prompt records are the bootstrap units and 5,000 draws use seed 20270916. Strict cycles require all three directed edge margins to be nonzero and consistently oriented. The

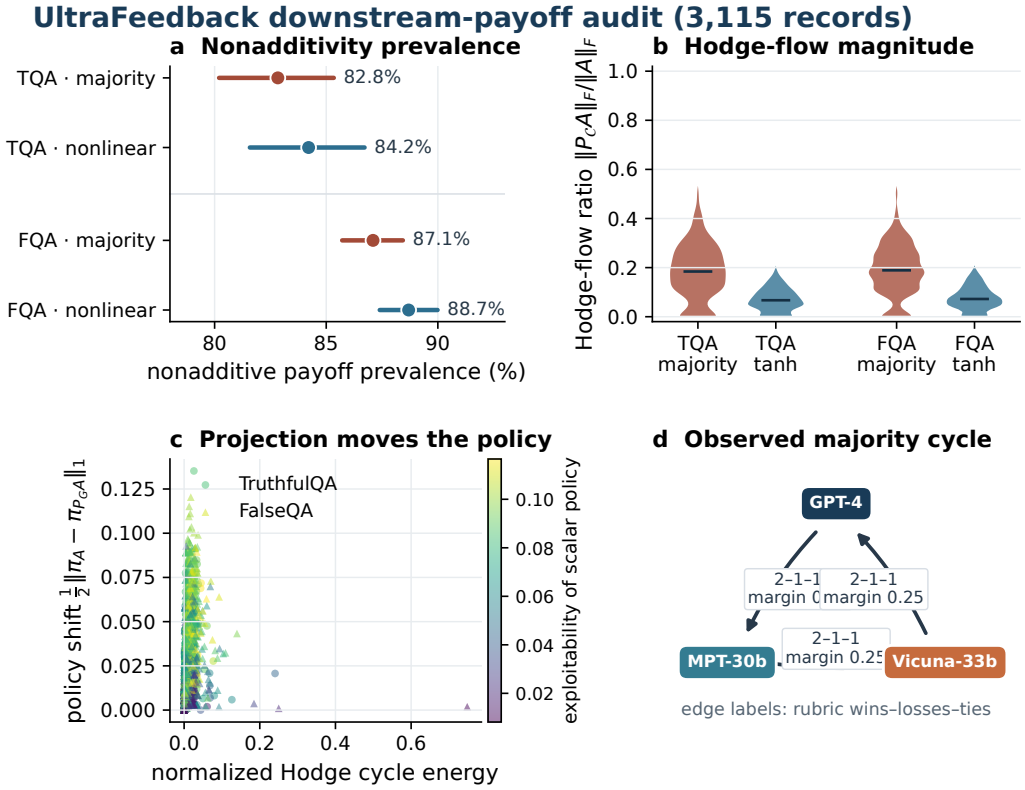


Figure 13: Downstream-payoff audit over 3,115 complete UltraFeedback records. Panel a gives prompt-bootstrap 95% intervals, panel b separates Hodge magnitude from strict sign cycles, panel c reports policy total variation at $\tau = .2$, and panel d shows the deterministically selected rubric-majority example.

representative cycle is selected by a deterministic rule that maximizes the minimum majority margin, with Hodge energy only as a secondary tie-breaker.

C.13 Implementation safeguards

The simulation globally rescales the truth, asserts component ranks and Hodge identities before sampling, uses a noncancelling registered cycle path, and solves the exact displayed constrained objective. Every plot exposes the registered metric range. The manifest records seeds and package versions; the generator avoids draws in non-unique singular-vector bases, making the generated problem invariant to arbitrary basis rotations.

C.14 Claim–evidence map

Table 7: Each empirical result targets a specific claim or failure mode; the deployment-geometry rows concern constructed downstream payoffs.

Claim under stress	Falsification design	Observed result
The pipeline preserves scalar negative controls	Generate $\Theta P_C = 0$ and assert projector identities	cycle RMS $< 2.0 \times 10^{-16}$; scalar KL and population-target policy TV zero
Finite-sample size, resolution, and heterogeneity	Apply Theorem 4.1 with 20,000 replications per condition	zero wrong resolutions; boundary always indeterminate; both separated branches resolve at least 99.7%

Claim under stress	Falsification design	Observed result
Rowwise auditing selects and certifies a hybrid representation	Audit five rubrics jointly, retain relational estimates for unresolved rows, and compare every method on the same complete query family	at $n = 4096$, zero wrong resolutions; hybrid/scalar payoff error $.0486/.3288$; source, payoff, and policy certificate coverage one across all 1,800 query instances
The rowwise hybrid solves the finite source-risk problem	Enumerate all four two-row actions, sweep cycle bias, and compare both uniform rules	Exhaustive and closed-form optima agree; Monte Carlo matches exact risk, with gains of 36.4% over all-scalar and 12.5% over all-relational
The finite-sample audit approaches the constrained oracle	Split audit from recovery, activate both policy- and risk-limited gates, and test (4.9)	190,207 covered events; zero unsafe projections or remainder violations; excess risk zero by $n = 1024$
The complete audit–recover–select–decide chain lowers risk with zero observed unsafe projection	Pair exact Bernoulli audit and independent recovery across six baselines, ablations, sensitivities, and 96 replications	weighted source risk .6119; 38.9%/23.4% below forced scalar/relational; zero unsafe projection; 76.2% certified resolution
The explicit recovery envelope covers its target event	Use $\lambda_j = 2u_j$ and cross the stated full-coverage threshold	all 768 above-threshold joint-event runs are covered; high-budget error/envelope slopes are $-.488/-.505$
Representation has an efficiency–bias crossover	Match likelihood, solver, calibration, and tolerance	at $c = 0$, scalar/Hodge RMSE .206/.244; at $c = 1$, .430/.312
Hodge sharing improves over a modern relational recovery baseline	Run the pinned official PCWST solver on the paired cycle–budget grid	Hodge lowers five estimation/decision losses in all nine conditions; every problem-level omnibus Holm $p < 1.7 \times 10^{-5}$
Sparse guarantees require design curvature or tolerance	Pair uniform, concentrated, and structurally missing sampling and compute registered tangent diagnostics	uniform curvature reaches .789; structural missingness retains zero curvature and at least 36.8% nullity
Exact-cell scalarity is resolved when coverage separates the confidence box from the gradient space	Repeat exact prompt–candidate cells, deduplicate workers, and invert simultaneous binomial intervals	At $\rho = 0$, 183 cells and 3,219 votes yield the indeterminate branch; localized Holm tests reject 0/30 prompts
The analytic audit tolerates worker-specific probabilities	Invert simultaneous non-identical-Bernoulli Hoeffding intervals for $\bar{p}_d$	all 30 prompt graphs remain indeterminate; radii .460–.667
Policy tolerance determines an operational source margin	Sweep $\varepsilon_\pi \in [0, 2]$ using $\rho = 2\tau\varepsilon_\pi$ at $\tau = .2$	0 adequate, 30 indeterminate, 0 nonintegrable throughout; smallest finite threshold 12.77
Sample circulation can affect a decision before population scalarity is resolved	Compare plug-in policies with and without the empirical Hodge component	median TV .026, 90th percentile .120 at $\tau = .2$
Deployment geometry depends on declared semantics	Compare linear logits, majority, and expected-win margins	linear cycle energy is zero; nonlinear payoffs are nonadditive in 82.8–88.7% of records
Magnitude nonadditivity can persist within one rubric	Remove each rubric, then audit each rubric alone	leave-one-out 71.8–88.1%; single-rubric nonzero-Hodge prevalence 32.6–72.7%, with strict-cycle rate zero

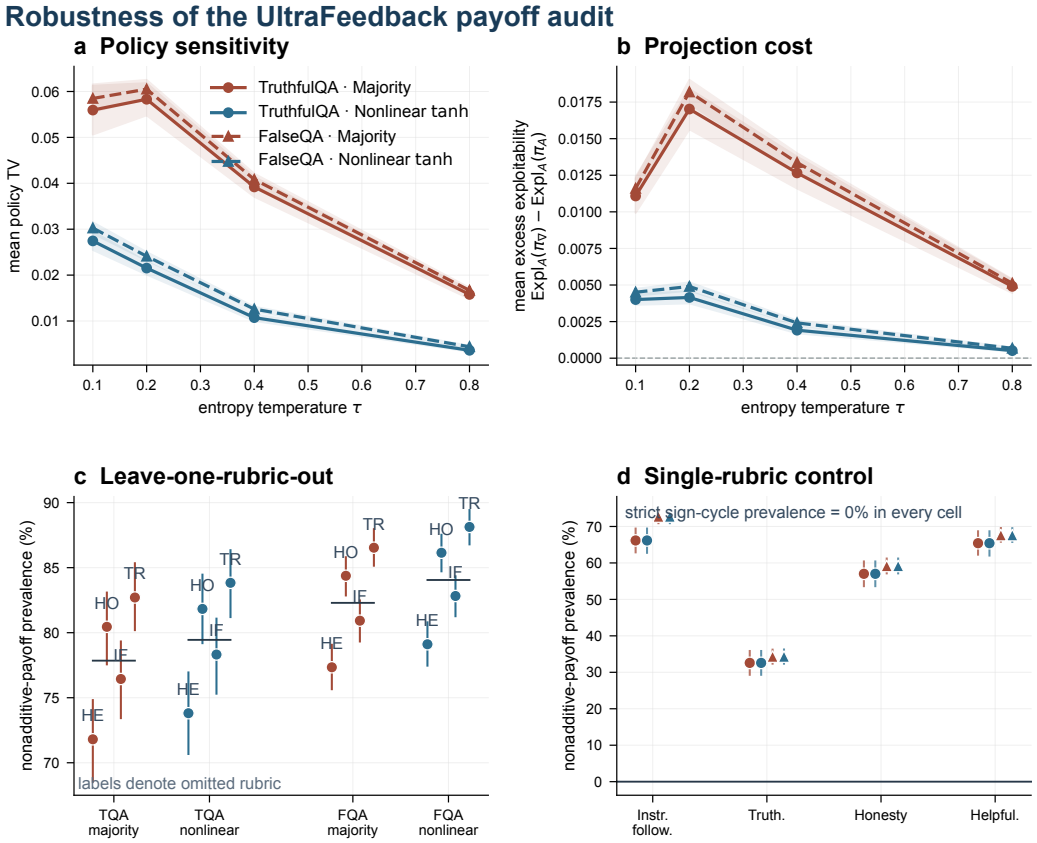


Figure 14: Payoff-audit robustness. Panel a gives policy sensitivity over $\tau = .1$–$.8$; expected-win sensitivity declines throughout, while majority sensitivity peaks near .2. Panel b reports the excess exploitability from gradient projection, $\mathrm{Expl}_A(\pi_{\mathrm{gradient}}) - \mathrm{Expl}_A(\pi_{\mathrm{full}})$. Panel c removes each rubric, and panel d audits each rubric alone. Points are record-level means; bands and whiskers are 95% prompt-bootstrap intervals. Single-rubric nonzero-Hodge prevalence is 32.6–72.7%, while every corresponding strict-cycle rate is zero.

D Lean 4 Verification Scope

The `LeanProofs` supplement is pinned to Lean `v4.34.0-rc1` and the matching mathlib revision. The command

```
cd LeanProofs
lake env lean LeanProofs/RubricGames.lean
lake build
```

kernel-checks the following core results:

1. triangle circulation of scalar rewards is zero;
2. a strict three-cycle lies outside the scalar-representable family;
3. scalar rewards telescope on finite paths and closed walks;
4. preserving every scalar triangle is equivalent to Cauchy additivity;
5. linear entrywise transforms preserve every scalar triangle;
6. every finite-dimensional flow has a unique gradient/orthogonal split;
7. the Hodge residual satisfies the Pythagoras identity;

8. the cycle norm is the attained scalar floor, with an exact excess-risk identity;
9. rowwise and rubric-weighted hybrid projection errors compose additively;
10. the rowwise threshold action globally minimizes every finite separable hybrid risk and attains its exact closed-form value;
11. weighted RMS parameter error controls a Lipschitz payoff aggregate;
12. on a three-response graph, cycle energy is squared circulation divided by three.

The source contains zero `sorry`, `admit`, unsafe definitions, or custom axioms. `#print axioms` reports the standard logical principles `propext`, `Classical.choice`, and `Quot.sound` inherited from Lean/mathlib. Lean verification covers the deterministic algebra behind Theorems 4.3 and 4.4; the preceding sections prove the expectation, continuity-to-linearity, concentration, restricted-curvature recovery, and Nash-stability results.