DEPLOYRANK: RANKING AMBIGUITY AND CERTIFIED CHOICE UNDER MISSING CONTEXT LINKAGE

Anonymous authors
Paper under double-blind review

ABSTRACT

Benchmarks may publish outcomes and all candidates' predictions for each evaluation record, yet omit the records' links to context available at deployment. We study what these releases identify and what evidence justifies model choice. For any finite number of candidates, we construct one complete-joint prediction release that preserves a strict standalone Blackwell order yet admits every contextual Blackwell order arbitrarily close to a fixed redundancy completion. Each requested order holds in one compatible world simultaneously for all regular strictly proper losses. Ranking ambiguity, however, need not prevent an ϵ-optimal choice. For frozen rules and binary context, DeployRank exactly tests zero-link decision adequacy and certifies choices through selectively disclosed row–context links. We derive exact pairwise minimum-disclosure frontiers and multi-candidate bounds, distinguishing holder-known proof size from online discovery cost. In a post-hoc minimax analysis of one protocol-frozen UCI Heart task, the release-only choice incurs .00790 nats of aligned regret at $\epsilon = .005$. An alignment-informed holder certifies the corrected choice with 86 of 417 links; the precommitted policy uses 278 paid links, counting quota-implied labels as free. Other audits demonstrate zero-link sufficiency and guarantee-only repair.

1 INTRODUCTION

Suppose an evaluator must choose a model for deployment across hospitals. The benchmark publishes every candidate's prediction and outcome for each evaluation record, plus outcome–hospital counts, but withholds which hospital each prediction row came from. Hospital identity *will* be available at deployment; it is the historical prediction–hospital linkage that is missing from evaluation. Does the public release already justify a choice, or must the evaluator obtain some of those links? The distinction matters: choosing poorly can incur deployment loss, but reconstructing a full ranking can require evidence that the decision never needed. Our central insight is to *test whether a tolerable choice is already certified before acquiring more alignment*. When an evaluator trusts the holder's aligned risks, those risks can settle selection directly. Our contract instead requires a guarantee that another evaluator can replay from public losses, quotas, and authenticated links.

The release retains the *complete* joint law of all candidate predictions and outcomes, $P(Y, Z_{1:M})$, as well as $P(Y, C)$. It therefore preserves candidate–candidate dependence, not merely a separate performance summary for each model. What remains free is their common alignment with C. We call the joint laws reproducing both tables the *release fiber*. Conditional value is familiar: under log loss, standalone and context-assisted values are $I(Y; Z_m)$ and $I(Y; Z_m \mid C)$, and their rankings may differ (Hewitt et al., 2021; Bertschinger et al., 2014). Our question is stronger: how many contextual rankings can *one fixed release* support? The difficulty is shared feasibility: all candidates must use one context alignment while both published laws stay fixed. Separate candidate intervals can combine incompatible worlds; choosing a conditional-independence coupling instead assumes the missing information rather than certifying a choice against it.

A complete ranking is more than a deployment decision requires. DeployRank first tests whether one frozen rule is within ϵ of the best under *every* compatible alignment. If so, it certifies that choice without requesting any linkage, even when the ranking remains ambiguous. Otherwise, selected links can exclude worlds in which the proposed choice fails.

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.

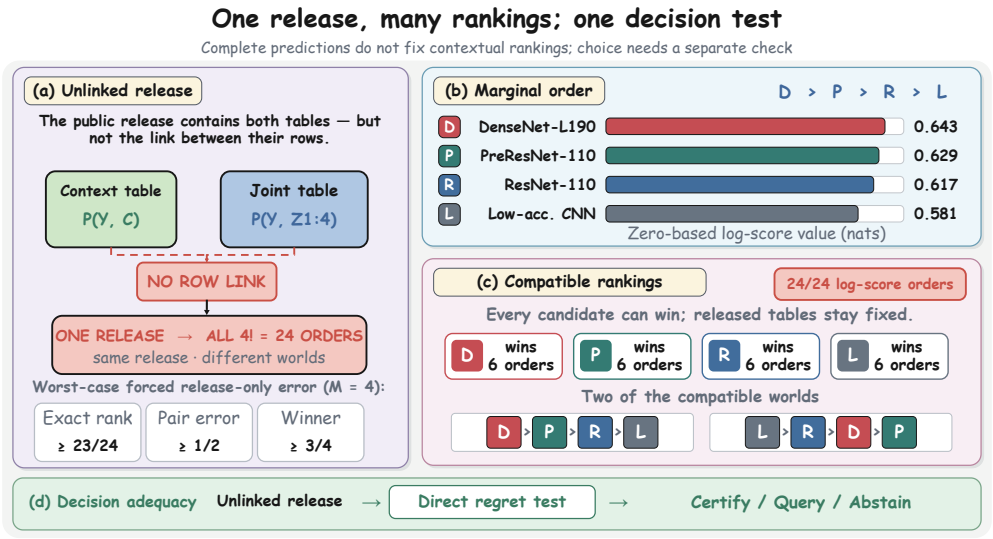

Figure 1: **One unlinked release, all rankings, and a separate decision test.** Panels (a)–(c) show one CIFAR-10H release fixing $P(Y, C)$ and complete $P(Y, Z_{1:4})$, a strict marginal log-score order, and all 24 compatible context-assisted log-score orders (zero count error; total variation (TV) at most 0.01014 from the conditional-independence completion; post-hoc exact-row witnesses, Appendix D.12). Cards give worst-world minimax error lower bounds for any forced release-only ranking, including randomized rules. Panel (d) depicts the general fixed-rule adequacy test: ranking ambiguity alone does not imply a need for disclosure; zero-link certification depends on whether $\min_a W_a(\varnothing) \leq \epsilon$. The panel does not claim a CIFAR zero-link certificate. Brier/spherical realize 13/24 and 11/24 orders.

This gives a three-step evaluation route: specify the release constraints, test zero-link adequacy, and disclose selected links only when needed. The shortest holder-known proof and the queries needed to find it are separate costs. Our contributions are:

Release-level non-identification. We construct a single complete-joint release supporting every contextual Blackwell order locally and simultaneously across regular strictly proper losses. For finite fixed rules, we characterize when the compatible risk set supports every order.

Decision-directed certification. For binary context, we derive a sharp zero-linkage criterion for fixed-rule ϵ-optimality and an exact pairwise disclosure frontier. Multi-candidate bounds and checkable exactness conditions yield publicly verifiable selection certificates.

Empirical separation of ambiguity, repair, and discovery. Decision audits separate zero-link sufficiency from repair. One UCI Heart task repairs an above-tolerance choice error and compares minimum proof with discovery.

2 Related Work

Conditional value and proper-score comparison. Added-value analysis, conditional probing, proper scores, and Blackwell comparison formalize what a predictor contributes beyond a baseline once the relevant joint law is observed (Boulesteix & Sauerbrei, 2011; Moons et al., 2012; Hewitt et al., 2021; Yang et al., 2020; DeGroot & Fienberg, 1983; Blackwell, 1953; Gneiting & Raftery, 2007). We study the preceding identification question when that joint law is incomplete: one fixed release must admit the claimed ranking in a common completion for all candidates. Candidate-wise value intervals do not establish this shared feasibility.

Partial identification, matching, and data fusion. Statistical matching, partial information decomposition (PID), contingency-table fibers, and incomplete multimodal fusion expose dependence left free by observed margins (Cross & Manski, 2002; Ahfock et al., 2016; Bertschinger et al., 2014;

Slavković et al., 2014; Jiang & Janson, 2025; Koo et al., 2026; Liang et al., 2024). Semiparametric data-fusion methods estimate endpoints of moment-based outer bounds from sampled data, while completion-selection methods add assumptions to choose a coupling. We condition on fixed published laws and compute sharp guarantees over their compatibility fiber. Estimator efficiency and exact release-level sharpness therefore answer complementary questions.

Decision adequacy and active information acquisition. Deployment-complete benchmarking studies decision adequacy over an evidence fiber; decision-region methods such as Hyperedge Cutting (HEC) query until the remaining hypotheses justify a choice (Mansouri & Arai, 2026; Fitas, 2026; Javdani et al., 2014). For quota-constrained row linkage, we characterize the exact proof that makes a choice publicly checkable, separately from the online cost of finding it. Appendix F.1 develops count-state HEC and minimax planning quotients, with their distinct prior-dependent and worst-case guarantees.

Structured disclosure and uncertain rankings. Minimum sufficient reasons already fix coordinates until every allowed completion preserves a verdict; sorting is exact for an unconstrained hypercube, and general feasible subsets encode domain restrictions (Barceló et al., 2020; Gorji & Rubin, 2021). Here exact released quotas impose a cardinality constraint within each outcome stratum, while one shared row set must satisfy all challenger inequalities. We derive the resulting prefix frontier, cross-stratum composition, and joint-candidate bounds. Matching interdiction (Zenklusen, 2008) concerns related leader–follower hardness, but not this quota-constrained certificate. Sampling-based rank confidence sets address uncertainty in estimated risks (Hansen et al., 2011; Neuhof & Benjamini, 2024; 2026): our released laws are fixed; uncertainty lies in their row alignment. Appendix I compares these information contracts and guarantee types directly.

3 What an Unlinked Release Identifies

3.1 Release and evaluation objects

Information contract. Public: each row's outcome and complete candidate-prediction vector, plus outcome–context counts. Hidden: the true row–context link, although context is available at deployment. For a public frozen rule A_m, the evaluator can compute $\ell_m(i, c) = \ell(y_i, A_m(z_{im}, c))$ for every possible context; only which entry applies to the row is hidden. Output: a candidate and worst-case regret bound replayable from public inputs and disclosed links. This guarantee concerns the fixed table and frozen rules, not future samples or retraining.

Three-row preview. One outcome stratum has three rows and exactly one context-one label. For two frozen rules a, j, the public loss differences are $(0, -2, 0)$ under context zero and $(3, 0, -2)$ under context one. The three compatible assignments give mean differences $1/3$, 0, and $-4/3$; the true assignment is $(0, 1, 0)$. Against this challenger, a needs no links at $\epsilon \geq 1/3$. At $\epsilon = 0$, revealing row 1's true zero excludes the adverse assignment and certifies a; revealing row 3's zero does not. Useful links block adverse assignments, rather than merely reducing unknown rows.

For population Bayes acts, let $U^\star_m = \mathcal{R}^*(\varnothing) - \mathcal{R}^*(Z_m)$ be standalone value and $V^\star_m(C) = \mathcal{R}^*(C) - \mathcal{R}^*(C, Z_m)$ be context-assisted incremental value; larger is better. For a frozen rule, $R_m(Q)$ is its mean loss; smaller is better. A *regular strictly proper loss* on finite $\mathcal{Y}$ has continuous, strictly concave Bayes risk $H_\ell(p) = \inf_a \mathbb{E}_{Y\sim p}\ell(Y, a)$, and for finite signals $\mathcal{R}^*_\ell(X) = \sum_x P(x) H_\ell(P(Y = \cdot \mid x))$. Table 1 separates Bayes-optimal signal use from frozen-rule certification, whose risks are affine in alignment.

The core release $\widetilde{\mathcal{O}} = \{P(Y, C), P(Y, Z_{1:M})\}$ preserves outcome–context statistics and all outcome–candidate and candidate–candidate dependence, but omits linkage to C. Write $\mathcal{C}(\mathcal{R})$ for the full joint laws reproducing release $\mathcal{R}$. Appendix D treats the weaker candidate-wise release.

3.2 One release, every contextual order

The release, reference completion, and neighborhood scale below are fixed *before* the requested contextual order is chosen.

Table 1: **Evaluation objects and the guarantees available for each.** Each row states what is analyzed and its principal boundary.

Object	Quantity	Available conclusion	Does not imply
Bayes signal value	$V_m^*(C)$	identified intervals; tie/tangent test; local all-$M!$ shattering	fixed-rule or fitted-pipeline behavior
Frozen fixed rule	$R_m(Q)$	all-order criterion; sharp zero-alignment boundary; regret and row certificates	Bayes-optimal use or refitting

Theorem 1 (One fixed release shatters the contextual Blackwell order) *For every finite $M \geq 2$, there exist binary $Y, C, Z_1, \ldots, Z_M$, one release $\widetilde{\mathcal{O}} = \{P(Y, C), P(Y, Z_{1:M})\}$, a full-support completion $Q_0 \in \mathcal{C}(\widetilde{\mathcal{O}})$, and $\eta_0 > 0$ with the following properties. First, the standalone experiments are strictly Blackwell ordered, $Z_1 \succ_B \cdots \succ_B Z_M$. Second, at Q_0 every Z_m is redundant given C: $Y \perp Z_m \mid C$, simultaneously for all m. Third, for every $0 < \eta < \eta_0$ and every permutation π, one completion $Q_{\pi,\eta} \in \mathcal{C}(\widetilde{\mathcal{O}}) \cap B_{\mathrm{TV}}(Q_0, \eta)$ satisfies the strict contextual experiment order*

$$(C, Z_{\pi(1)}) \succ_B (C, Z_{\pi(2)}) \succ_B \cdots \succ_B (C, Z_{\pi(M)}).$$

Here $\succ_B$ means more informative: it can simulate the weaker experiment by outcome-independent randomization, but not conversely. Consequently, the same *$Q_{\pi,\eta}$ realizes $V^*_{\pi(1),\ell}(C) > \cdots > V^*_{\pi(M),\ell}(C)$ simultaneously for every regular strictly proper loss ℓ; at Q_0 all these incremental values are zero.*

Mechanism and scope. Let E_m be independent binary symmetric-channel errors with nearby distinct rates q_m, and let F be a binary context error. Set $Z_m = Y \oplus E_m$ and $C = Y \oplus F$: the release fixes the signal-error law, while the missing link permits changes in how F depends on the E_m. The construction factorizes the joint law as

$$P(F = 1 \mid E_{1:M}) = r + \sum_m \theta_m (E_m - q_m).$$

Centering keeps both published laws fixed; changing θ_m adjusts only candidate m's pairwise context-error link. At a fixed redundancy vector $\theta^\star$, the posterior intervals of the context-error state F given $W_m = C \oplus Z_m$ collapse to the common prior r; geometrically separated perturbations nest them in any requested order and yield outcome-independent Blackwell garblings. Thus no contextual winner, nontrivial top-k set, or pairwise order is identified on this release.

The theorem does not assert a numerical gap uniform over losses: smooth-score gaps vanish near redundancy. A separate fixed-loss construction is linear in TV under a nonzero-derivative condition (Theorem 5). Exact rational audit: Appendix A.2; tangent-space and calibrated-forecast extensions: Appendix D.

We now move from the best possible use of each signal to rules fixed before evaluation. All candidates share the same unknown alignment; their joint risk vector, rather than separate candidate intervals, determines attainable rankings.

Theorem 2 (Global-to-local characterization for finite fixed rules) *Let Γ be a nonempty convex polytope of compatible finite-alphabet completions, and let $R\colon \operatorname{aff}\Gamma \to \mathbb{R}^M$ be the affine risk vector of $M \geq 2$ fixed rules. Let $\mathcal{K}_\Gamma = \{R(Q) - \overline{R}(Q)\mathbf{1} : Q \in \Gamma\}$, where $\overline{R} = M^{-1}\sum_m R_m$. Every strict ranking is compatible exactly when $\mathcal{K}_\Gamma$ contains a neighborhood of zero in $\mathbf{1}^\perp$. Equivalently, some $Q_0 \in \operatorname{ri}\Gamma$ ties all risks and the centered risk map has rank $M - 1$ on the affine directions of Γ. Whenever these conditions hold, every strict order occurs in every relative neighborhood of the same Q_0.*

Subtracting the mean risk removes shifts that change no ranking. The remaining risk vectors support every order precisely when they surround a tie; full rank without an interior tie is insufficient. Appendix A gives the equivalent derivative formulation and proof. This characterization concerns the convex completion polytope; at fixed empirical sample size, strict-order realizability additionally

requires an integer witness. If every order is attainable with adjacent risk gap at least g, every fixed choice has worst-compatible regret at least $(M-1)g$ (Corollary 3). Order attainability alone gives no loss gap above a specified ϵ. The general margin corollary is in Appendix A; artifact-specific certificates are in Appendix D.12.

4 Certifying a Choice with Selective Links

The all-order result concerns rankings; a deployment decision asks less. For a transcript S of disclosed row–context links, let $\Gamma(S)$ be its compatible completions and define a candidate's worst-compatible regret

$$W_a(S) = \max_{j\in[M]} \Phi_{aj}(S), \qquad \Phi_{aj}(S) = \sup_{Q\in\Gamma(S)} \{R_a(Q) - R_j(Q)\}. \tag{1}$$

The release and answers certify a choice exactly when $\min_a W_a(S) \le \epsilon$. If this fails, the evaluator must abstain or request another link. The finite maximum and supremum commute: different challengers may have different worst worlds because safety requires the same chosen action to withstand *every* compatible world, not a joint ranking witnessed by one world. Here "exact" means sharp for the frozen finite loss table and quota fiber; "ϵ-safe" means guaranteed under that contract, not clinically safe. An aligned error is material only when regret exceeds the prespecified ϵ.

Algorithm 1: Replayable choice certification.
Input: frozen loss table, outcomes, public context quotas, tolerance ϵ, a release-visible query policy π, and paid-query budget $q_{\max}$.
1. Initialize $S_{\text{paid}} = \varnothing$ and close answers under the quotas: $S = \text{cl}(S_{\text{paid}})$.
2. Compute every $W_a(S)$ and choose a minimizer a. If $W_a(S) \le \epsilon$, return a, S_{paid}, and the certified bound.
3. If $|S_{\text{paid}}| = q_{\max}$, return UNRESOLVED. Otherwise query an unforced row using π and prior answers, authenticate its answer, append it to S_{paid}, close the quotas, and return to step 2.
A contradictory answer fails the audit. Quota-implied labels enter the verifier but do not count as paid queries.

The query policy determines *how quickly* this test passes; the exact stopping test determines whether its returned choice is valid.

4.1 When no linkage is needed

For deterministic selection from the declared finite rule menu, the decision relevance of ranking ambiguity is measured by release-level minimax regret

$$\mathfrak{G}(\Gamma) = \min_{a\in[M]} \sup_{Q\in\Gamma} \left\{ R_a(Q) - \min_{j\in[M]} R_j(Q) \right\}. \tag{2}$$

Thus $\mathfrak{G}(\Gamma) \le \epsilon$ exactly when some candidate is ϵ-optimal under every compatible completion; otherwise no release-only fixed choice has that guarantee. Theorem 3 computes this boundary for binary context.

Theorem 3 (Sharp decision adequacy from an unlinked release) *Let $C \in \{0,1\}$, $X = (Y, Z_{1:M})$, and let the release fix P_X and $P_{Y,C}$ but not their alignment. For fixed losses in $[0,B]$, define $b_{aj} = \ell_a(X,0) - \ell_j(X,0)$, $\delta_{aj} = \ell_a(X,1) - \ell_j(X,1) - b_{aj}$, and $p_y = P(C=1 \mid Y=y)$. Then*

$$\Phi_{aj} := \sup_Q (R_a - R_j) = Eb_{aj} + \sum_y P(y)\mathcal{T}_{p_y}(\delta_{aj} \mid y) = \inf_\lambda E[b_{aj} + (\delta_{aj} - \lambda(Y))_+ + C\lambda(Y)],$$

where $\mathcal{T}_p(V) = \sup_{0\le q\le 1,\, Eq=p} E(qV)$. Thus $\mathfrak{G} = \min_a \max_j \Phi_{aj}$ is a sharp zero-alignment criterion; on an equal-weight finite table it is computed by stratum-wise top-k sums.

This sharp upper-tail functional aligns context with the largest loss increments within each outcome stratum. Equivalently, each Φ_{aj} is its contrast under the public conditional-independence completion plus a nonnegative *alignment premium*. A candidate needs no linkage precisely when its conditional-independence contrast plus that premium is at most ϵ against every challenger. Appendix B proves this decomposition and the dual formula.

4.2 WHAT THE EVALUATOR VERIFIES

We price exact revelation of a row's context link in a fixed finite table; outcomes, candidate predictions, and outcome–context counts are public. Certification requires one candidate to be ϵ-optimal under every remaining assignment.

Fix the release, losses, and tolerance before disclosure. The holder returns S with authenticated answers c_i. For proposed candidate a, challenger j, and row i, put $d_i(c) = \ell_a(i, c) - \ell_j(i, c)$ and $\Delta_i = d_i(1) - d_i(0)$. Let N be the number of rows, K_y the released number of context-one rows in stratum y, $S_y = \{i \in S : y_i = y\}$, and $U_y = \{i \notin S : y_i = y\}$. Here TopSum_k sums the k largest entries. In the three-row preview, $\Delta = (3, 2, -2)$: the unknown context-one label is assigned to the largest remaining increment. For any transcript, the evaluator computes this worst remaining assignment from public inputs:

$$\Phi_{aj}(S) = \frac{1}{N}\left[\sum_{i\in S} d_i(c_i) + \sum_{i\notin S} d_i(0) + \sum_y \mathrm{TopSum}_{K_y - \sum_{i\in S_y} c_i}\{\Delta_i : i \in U_y\}\right]. \tag{3}$$

The evaluator need not receive or run the holder's full aligned risk calculation: it checks each answer against the agreed row identity and public quota, then replays this expression from the public losses and transcript. Authentication of the underlying records is an external trust assumption, not something the optimization can establish. With truthful authenticated answers, S is sufficient exactly when $\max_{j\neq a} \Phi_{aj}(S) \le \epsilon$, computable without undisclosed context. This verifies safety; a separate frontier audit certifies minimality when its lower and upper bounds meet. For an explicit binary N-row table, this fixed-transcript verifier uses stratum-wise sorting/top-k sums per challenger; it neither enumerates the $M!$ rankings nor searches all compatible alignments. The harder tasks are finding a shortest joint proof and planning an optimal adaptive query policy.

4.3 PROOF SIZE IS NOT DISCOVERY COST

Given the true alignment, the holder seeks the smallest set of links that excludes every assignment making its choice more than ϵ suboptimal. This proof size is $\kappa^{\mathrm{select}}_\epsilon(c^\star)$, a minimum hitting set (Proposition 1). The separate regret radius measures distance to a bad world, not proof size (Proposition 7).

Corollary 1 (Material ambiguity forces linear exact linkage) *For the full binary-context row-assignment fiber Γ of N equal-weight rows, fixed losses in $[0, B]$ with $B > 0$, and $G_0 = \mathfrak{G}(\Gamma)$, every terminating transcript of any adaptive zero-error policy certifying some candidate at tolerance ϵ reveals at least $\lceil N(G_0 - \epsilon)_+/(4B)\rceil$ links. A uniform loss bound and $G_0 - \epsilon \ge c > 0$ uniformly in N therefore require $\Omega(N)$ links.*

The bound follows because one reveal can contract a candidate's sharp worst compatible regret by at most $4B/N$; Appendix C.2 proves this pathwise statement and gives sharper heterogeneous row bounds.

Theorem 4 (Exact pairwise disclosure frontier) *Fix a true binary alignment $c^\star$ and use the contrasts d_i and increments Δ_i defined above. Within outcome stratum y, sort the Δ_i of true-zero rows as $z_{y,1} \ge \cdots \ge z_{y,n_{y0}}$ and those of true-one rows as $o_{y,1} \le \cdots \le o_{y,n_{y1}}$. Among all certificates revealing u true zeros and v true ones in this stratum, one optimum reveals the first u rows of z_y and the first v rows of o_y; its remaining adversarial gain over the truth is*

$$E_y(u, v) = \sum_{r=1}^{\min(n_{y0}-u,\, n_{y1}-v)} [z_{y,u+r} - o_{y,v+r}]_+.$$

Consequently the exact pairwise minimum is

$$\kappa_{a,j}(\epsilon) = \min_{\{u_y, v_y\}} \left\{ \sum_y (u_y + v_y) : R_a(c^\star) - R_j(c^\star) + N^{-1} \sum_y E_y(u_y, v_y) \le \epsilon \right\},$$

computable by sorting and min-plus dynamic programming in $O(N^2)$ time.

The preview's useful first-row answer illustrates the prefix rule: reveal links that block harmful quota-preserving reassignments. Appendix C.2 extends the binary frontier to fixed-category context

and locates the corresponding tractability boundary; Section 5 reports the post-hoc $K = 3$ COIL bracket. These are holder-side results with known alignment, not online discovery guarantees.

Corollary 2 (Joint-selection consequences) *Let $\mathcal{A}_\epsilon(c^\star) = \{a : R_a(c^\star) - \min_j R_j(c^\star) \le \epsilon\}$ be the truth-feasible candidates. For $a \in \mathcal{A}_\epsilon(c^\star)$, let $L_a = \max_{j \neq a} \kappa_{a,j}$. Fix one minimum certificate per challenger, and let U_a^{union} be the smallest jointly sufficient union of a subcollection of them. Then*

$$L = \min_{a \in \mathcal{A}_\epsilon(c^\star)} L_a \;\le\; \kappa_\epsilon^{\mathrm{select}}(c^\star) \;\le\; U_{\mathrm{union}} = \min_{a \in \mathcal{A}_\epsilon(c^\star)} U_a^{\mathrm{union}} \;\le\; (M-1)L.$$

Equality holds when one fixed size-L pairwise certificate protects an $L_a = L$ candidate against every challenger. Shared within-stratum challenger orders yield an exact count-grid; Appendix C gives the strict joint-versus-union example with pairwise sizes $2, 2$, union 4, and joint optimum 3.

Without shared orders, threshold decision is NP-complete when M is part of the input, even for one stratum and a unique true winner; the unrestricted problem still has an exact mixed-integer linear program (MILP). Appendix C.2 details the shared-order grid, and Proposition 2 and Theorem 6 give the unrestricted formulation and hardness reduction.

Proof sketch. Within each stratum, quota-preserving reassignments pair unrevealed true-zero and true-one rows; a swap changes the contrast by $N^{-1}(\Delta_i - \Delta_{i'})$. Exchange gives the extreme prefixes, min-plus convolution allocates disclosure across strata, and pairwise minima lower-bound the joint certificate while their union suffices. The frontier uses native unbinned rows; Appendix C.2 gives the edge cases and joint-versus-union proof.

Discovery and computational scope. Proof size and zero-error discovery depth differ; query ratios are not competitive ratios. Appendix F.1's exact minimax and Bayesian HEC quotients can be exponential in public type count; empirical priorities have no competitive guarantee. By contrast, $O(M^2)$ linear programs (LPs) compute sharp risk bounds and regret on an explicitly represented finite completion polytope; the binary verifier uses top-k sums. Explicit release size can be exponential in M. These results exclude arbitrary refits; Bayes-optimal value remains nonlinear. The empirical appendix reports operating characteristics.

5 Empirical Evidence

Four questions organize the evidence: ranking compatibility, zero-link choice, holder proof size, and online discovery. Evaluators know predictions, outcomes, context quotas, and frozen losses, not row linkage. Quotas and prior answers force some labels for free. Policies share the exact stopping test; zero-query certification demonstrates release sufficiency, not policy superiority.

Question 1: Can one release hide different rankings? Equal-Y reallocations preserve both released laws and each row's complete candidate vector. Table S1 in the empirical appendix summarizes fixed-artifact witnesses, which test ranking identification rather than query-policy savings. Native rows evaluate post-hoc fixed rules; analysis chronology is recorded in Appendix S5.

The post-hoc, nonlocal CXR stress test certifies 120 Blackwell orders (minimum adjacent log-value gap .0374323; TV .3093–.3420). CIFAR certifies 24 log-score orders within TV .0102 of the observed alignment and every winner within .0056; Brier/spherical realize 13/24 and 11/24 orders. These are separate locality audits: Figure 1's .01014 uses the conditional-independence center Q_{CI}, whereas .0102 here uses the observed center Q_{obs}; the constructed theorem reference Q_0 is neither empirical center.

From ranking compatibility to decision regret. Native ImageNet-16H fixed rules realize 24/24 exact-row orders per score, with score-specific completions. On the 4,800-row log release, post-hoc minimax regret is .05026 nats for a 41-rule menu; on the same menu, a two-world randomized lower bound is .02956. The aligned choice is correct: these are worst-compatible regret bounds, not realized selection errors (Appendix D.12).

Question 2: Does the release already justify a choice? Three protocol-frozen releases certify the aligned winner without queries: UCI Diabetes has $G_0 = .002895 < \epsilon = .005$; PTB-XL and UCI Sepsis have $G_0 = 0$ (Section S2.8). Sufficiency has distinct sources. In audited Functional Map of the World (FMoW) settings, the rule stays optimal even without outcome–context quotas; in Bank and Online Shoppers, quotas exclude adverse assignments (Table S4).

Nonzero-linkage repair on UCI Heart. Before target access, we fixed three hospital pairs, four Cleveland-trained candidates, $\epsilon = .005$, and a Bernoulli Jensen–Shannon (JS) priority. Hospital identity is hidden. The frozen, non-refitted update uses public counts: $q_m(y \mid p, c) \propto p_m(y)P(c \mid Y = y)$. This is the audited rule, not a Bayes-posterior assumption across compatible alignments. A subsequent audit added release-only minimax choice and common free-quota accounting. On 417 H–S rows, $G_0 = .05684$; the release selects random forest (RF), whereas the aligned optimum is histogram gradient boosting (HGB). RF's .00790-nat aligned regret exceeds the prespecified $\epsilon = .005$. A holder can certify HGB with 86 links (20.6% of rows); frozen Bernoulli-JS discovers a certificate after 278 paid links (66.7%), versus 361.72 for uniform. The overlapping H–VA and S–VA settings repair a guarantee, not the action; their protocol and the CI-reference error are detailed in the empirical appendix (Table S10); the replay entry point is in the appendix guide (p. 13).

Table 2: **Decision consequence of missing linkage ($\epsilon = .005$).** G_0 is release-only minimax regret; Choice compares its action with the aligned optimum (= means unchanged). P denotes a frozen protocol; H denotes post-hoc minimax accounting. Zero-link tasks were protocol-frozen. Status and chronology: Table S3.

Setting	N	G_0	Choice	Aligned regret	Effect / status
Heart H–S	417	.05684	RF → HGB	.00790	choice / P+H
UCI Diabetes	4,096	.002895	ExtraTrees (=)	0	zero-link / P
PTB-XL	1,904	0	HGB (=)	0	zero-link / P
Waterbirds val	1,199	.24276	CLIP-L/14 (=)	0	guarantee / H
COIL-2000	4,000	.00676	Logistic (=)	0	guarantee / P+H
CelebA	5,000	.04098	EfficientNet-B0 (=)	0	guarantee / P+H

Question 3: What is the shortest holder-known proof? Table 3 pairs the global pairwise lower bound L with a separately verified sufficient set of size U_{cert}, not necessarily the union bound U_{union}. Equality proves minimality. On COIL, Waterbirds, and Heart, a minimum set against the hardest challenger also protects against all others. CelebA contrasts: its 263-link bottleneck is not jointly sufficient, leaving the verified 263–302 bracket. The $K = 3$ case uses Appendix C.2's categorical frontier. For Heart H–S, HGB alone is truth-feasible at $\epsilon = .005$; its pairwise minima are 35, 86, and 64 links, and the 86-link bottleneck set protects it against all challengers.

Table 3: **Proof size versus discovery at $\epsilon = .005$.** Exact-dyadic $[L, U_{\mathrm{cert}}]$ bounds holder proof size; equality proves the minimum (numerical contract: Appendix S2.9). CelebA's 263-link bottleneck is not jointly sufficient (upper bound 302). All query counts use common quota closure. EG/SR/JS denote expected-gap, selection-risk, and Bernoulli–JS. Other queries are frozen-secondary (S) or post-hoc (H); Waterbirds has no prospective primary.

Target	N	Holder proof	Frozen primary	Uniform	Other queries
COIL-2000 ($K = 2$)	4,000	$[159, 159]$	EG 336.62	911.04	SR 335.81 (S)
COIL-2000 ($K = 3$)	4,000	$[71, 238]$	—	—	—
Waterbirds val	1,199	$[123, 123]$	—	606.96	EG 184 (H)
Waterbirds test	5,794	$[619, 619]$	—	2,825.81	EG 995 (H)
CelebA	5,000	$[263, 302]$	EG 1,191.06	3,842.56	JS 1,019 (S)
Heart H–S	417	$\mathbf{[86, 86]}$	**JS 278**	361.72	—
Heart H–VA	494	$[71, 71]$	JS 212	327.54	—
Heart S–VA	323	$[69, 69]$	JS 151	239.58	—

Question 4: How many queries discover a proof? Online discovery can require substantially more links than the shortest holder-known proof. Table 3 reports predeclared primary policies separately from secondary and post-hoc comparisons.

At $\epsilon = .005$, frozen primary expected-gap saves 63.1%/69.0% versus uniform on COIL/CelebA. Neither prespecified superiority margin over informed baselines passes. The displayed secondary SR/JS policies save 63.1%/73.5%, selected post hoc by mean cost. I-CARE primary JS saves 23.5% and 17.0% in its two eligible settings, missing the requirement of at least 20% in every eligible setting (Appendix S2.7).

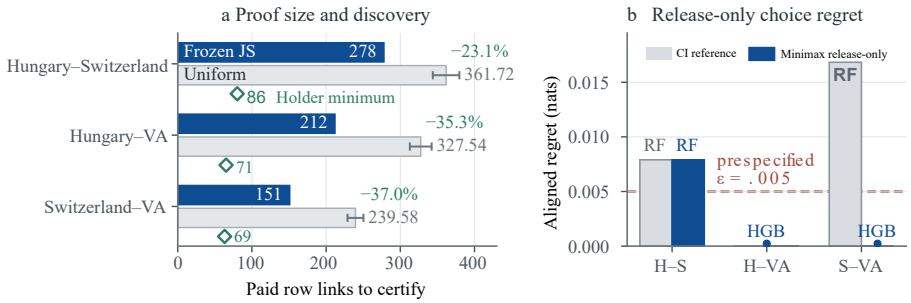

Figure 2: **Minimum proof size versus online discovery.** (a) Paid links under common quota closure; deterministic JS versus Uniform mean ± sample SD over 100 frozen permutations. JS paths coincide (no boundary ties). Diamonds mark separate holder minima (86/71/69), not an online policy. (b) CI versus minimax release-only regret; hospital pairs overlap. The 86-link H–S holder proof is a separate informed set, not a prefix of the 278-query JS path.

Marginal change is not yet release deficiency. Both 16-component principal component analysis (PCA-16) skin marginal winners change (excess losses .0111/.1308 nats; Tschandl et al., 2018; Combalia et al., 2019; Pacheco et al., 2020), yet both releases certify BiomedCLIP ($G_0 = 0$); these are not prevalence estimates (Table 8). Among primary decision settings, Heart H–S is the only completed repair of above-tolerance minimax choice error. In post-hoc 4/8/9-rule menus, fused RF still incurs .00790 regret against fused HGB, whereas adding a public-count context-only rule makes S–VA zero-link sufficient. Linkage need depends on both release and rule menu; 86 links and 278 queries apply only to the original four-rule menu (Table S12).

6 Discussion

Decision sufficiency before ranking resolution. The results separate three outcomes: a release already justifies a choice, linkage strengthens its guarantee without changing the winner, or linkage repairs a materially wrong choice. Ranking uncertainty cannot substitute for decision risk. Evaluation should first ask whether a candidate meets the specified tolerance across compatible alignments, rather than automatically requesting enough information to reconstruct the leaderboard. Guarantee-only repair explains why certification can require links even when the action stays unchanged. These audits establish distinct regimes, not their frequency.

Evidence under an explicit trust contract. In an illustrative audit, a benchmark operator releases frozen losses and quotas, an institution holds row-level context, and a separate evaluator must justify the choice. The institution attests selected row links without releasing its full alignment. An aggregate risk vector would settle selection if trusted, but does not let this evaluator recompute the guarantee from the released inputs. A link certificate supplies that replayable evidence. This is not a deployed external-holder pilot; authentication assumes truthful records rather than proving their correctness. Link count measures post-linkage release, excluding context collection, initial linkage, authentication, and governance costs.

From shortest proof to usable acquisition. Knowing which links suffice and discovering them without the true alignment are different optimization problems. Exact verification supplies a stopping rule and a target for acquisition; it does not make every query policy efficient. The next application step is independent-holder execution under the same contract, coupled with sampling or stability assumptions when the target extends beyond the frozen table to future observations or refitted models.

7 Conclusion

A release can admit every contextual order while still supporting a certified choice. Decision adequacy asks whether one frozen candidate stays within tolerance in every compatible world. Otherwise, sufficient evidence and its discovery cost are separate objects. DeployRank makes these questions checkable under an explicit release and evaluation contract.

Ethics statement

This work evaluates existing datasets and model outputs; it does not introduce a prospective patient study or clinical intervention. Medical and demographic contexts require particular care: certification under a fixed release does not establish clinical safety, fairness, or a privacy guarantee. We do not redistribute restricted patient-level artifacts or private row contexts. Public replay materials omit original clinical features and patient identifiers; the availability ledger distinguishes these materials from restricted inputs. Selected-link minimization reduces disclosure under the stated contract but does not by itself establish anonymity or prevent inference from released statistics. Dataset licenses and access conditions remain applicable.

Reproducibility statement

The proof appendices provide assumptions, complete proofs, and computational formulations; the empirical appendix records evaluation protocols, frozen policies, accounting conventions, and deviations. Its replay-scope table (Table S22) identifies which results can be checked from the anonymous archive, which require reacquiring public data, and which depend on restricted inputs. The archive's `examples/heart_replay/` and `examples/aligned_simulation/` provide public certificate and simulation checks. Aggregate records and hashes identify retained artifacts but do not replace unavailable inputs or permit reconstruction of private alignments.

References

Daniel Ahfock, Saumyadipta Pyne, Sharon X. Lee, and Geoffrey J. McLachlan. Partial identification in the statistical matching problem. *Computational Statistics & Data Analysis*, 104: 79–90, 2016. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2016.06.005. URL https://www.sciencedirect.com/science/article/pii/S0167947316301426.

Edilberto Amorim, Wei-Long Zheng, Jong Woo Lee, Susan Herman, Mohammad Ghassemi, Adithya Sivaraju, Nicolas Gaspard, Jeannette Hofmeijer, Michel J A M van Putten, Matthew Reyna, Gari Clifford, and Brandon Westover. I-CARE: International Cardiac Arrest REsearch consortium Database. *PhysioNet*, December 2023. doi: 10.13026/m33r-bj81. URL https://doi.org/10.13026/m33r-bj81. Version 2.1.

Pablo Barceló, Mikaël Monet, Jorge Pérez, and Bernardo Subercaseaux. Model interpretability through the lens of computational complexity, 2020. URL https://arxiv.org/abs/2010.12265.

Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, April 2014. ISSN 1099-4300. doi: 10.3390/e16042161. URL http://dx.doi.org/10.3390/e16042161.

Jock Blackard. Covertype. UCI Machine Learning Repository, 1998.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24 (2):265–272, 1953. ISSN 00034851. URL http://www.jstor.org/stable/2236332.

Daniel Borkan, Lucas Dixon, Jeffrey Sorensen, Nithum Thain, and Lucy Vasserman. Nuanced metrics for measuring unintended bias with real data for text classification, 2019. URL https://arxiv.org/abs/1903.04561.

Anne-Laure Boulesteix and Willi Sauerbrei. Added predictive value of high-throughput molecular data to clinical data and its validation. *Briefings in Bioinformatics*, 12(3):215–229, 2011. doi: 10.1093/bib/bbq085.

Davide Chicco and Giuseppe Jurman. Survival prediction of patients with sepsis from age, sex, and septic episode number alone. *Scientific reports*, 10(1):17156, 2020.

Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Cristina Carrera, Alicia Barreiro, Allan C. Halpern, Susana Puig, and Josep Malvehy. Bcn20000: Dermoscopic lesions in the wild, 2019. URL https://arxiv.org/abs/1908.02288.

Philip J. Cross and Charles F. Manski. Regressions, short and long. *Econometrica*, 70(1):357–368, 2002. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2692173.

Morris H. DeGroot and Stephen E. Fienberg. The comparison and evaluation of forecasters. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 32(1):12–22, 1983. ISSN 00390526, 14679884. URL http://www.jstor.org/stable/2987588.

Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. Retiring adult: New datasets for fair machine learning, 2022. URL https://arxiv.org/abs/2108.04884.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions, 2016. URL https://arxiv.org/abs/1404.3763.

Ricardo Fitas. The geometry of AI validation: From structural blindness to reusable audits. 2026. doi: 10.48550/arXiv.2608.21496.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437. URL https://doi.org/10.1198/016214506000001437.

Niku Gorji and Sasha Rubin. Sufficient reasons for classifier decisions in the presence of constraints, 2021. URL https://arxiv.org/abs/2105.06001.

Ziyang Guo, Yifan Wu, Jason Hartline, and Jessica Hullman. The value of information in human-ai decision-making. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 41826–41857, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/45824c7580778ad1ba384982b9e6b39f-Paper-Conference.pdf.

Peter R Hansen, Asger Lunde, and James M Nason. The model confidence set. *Econometrica*, 79(2): 453–497, 2011.

John Hewitt, Kawin Ethayarajh, Percy Liang, and Christopher D. Manning. Conditional probing: measuring usable information beyond a baseline, 2021. URL https://arxiv.org/abs/2109.09234.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs, 2021. URL https://arxiv.org/abs/2005.00687.

Andras Janosi, William Steinbrunn, Matthias Pfisterer, and Robert Detrano. Heart Disease. UCI Machine Learning Repository, 1989. DOI: https://doi.org/10.24432/C52P4X.

Shervin Javdani, Yuxin Chen, Amin Karbasi, Andreas Krause, J. Andrew Bagnell, and Siddhartha Srinivasa. Near optimal bayesian active learning for decision making, 2014. URL https://arxiv.org/abs/1402.5886.

Yicong Jiang and Lucas Janson. Semiparametric inference for partially identifiable data fusion estimands via double machine learning, 2025. URL https://arxiv.org/abs/2502.05319.

Richard M. Karp. *Reducibility among Combinatorial Problems*, pp. 85–103. Springer US, Boston, MA, 1972. ISBN 978-1-4684-2001-2. doi: 10.1007/978-1-4684-2001-2_9. URL https://doi.org/10.1007/978-1-4684-2001-2_9.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton A. Earnshaw, Imran S. Haque, Sara Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. Wilds: A benchmark of in-the-wild distribution shifts, 2021. URL https://arxiv.org/abs/2012.07421.

Eunho Koo, Tongseok Lim, and Jinwon Sohn. Statistical matching via schrödinger bridge beyond conditional independence, 2026. URL https://arxiv.org/abs/2606.22770.

Paul Pu Liang, Chun Kai Ling, Yun Cheng, Alex Obolenskiy, Yudong Liu, Rohan Pandey, Alex Wilf, Louis-Philippe Morency, and Ruslan Salakhutdinov. Multimodal learning without labeled multimodal data: Guarantees and applications, 2024. URL https://arxiv.org/abs/2306.04539.

El Mustapha Mansouri and Keigo Arai. Deployment-complete benchmarking, 2026. URL https://arxiv.org/abs/2605.25997.

Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization, 2009. URL https://arxiv.org/abs/0907.3740.

Karel GM Moons, Joris AH de Groot, Kristian Linnet, Johannes B Reitsma, and Patrick MM Bossuyt. Quantifying the added value of a diagnostic test or marker. *Clinical chemistry*, 58(10):1408–1417, 2012.

S. Moró, P. Rita, and P. Cortez. Bank marketing. UCI Machine Learning Repository, 2014. DOI: https://doi.org/10.24432/C5K306.

Bitya Neuhof and Yuval Benjamini. Confident feature ranking, 2024. URL https://arxiv.org/abs/2307.15361.

Bitya Neuhof and Yuval Benjamini. Rank intervals for leaderboards: A hierarchical framework for model evaluation, 2026. URL https://arxiv.org/abs/2606.08679.

Andre G. C. Pacheco, Gustavo R. Lima, Amanda S. Salomão, Breno A. Krohling, Igor P. Biral, Gabriel G. de Angelo, Fábio C. R. Alves Jr, José G. M. Esgario, Alana C. Simora, Pedro B. C. Castro, Felipe B. Rodrigues, Patricia H. L. Frasson, Renato A. Krohling, Helder Knidel, Maria C. S. Santos, Rachel B. do Espírito Santo, Telma L. S. G. Macedo, Tania R. P. Canuto, and Luíz F. S. de Barros. Pad-ufes-20: a skin lesion dataset composed of patient data and clinical images collected from smartphones, 2020. URL https://arxiv.org/abs/2007.00478.

Dimitris N. Politis and Joseph P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4):2031–2050, 1994. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/2242497.

Alexander Postnikov. Permutohedra, associahedra, and beyond, 2005. URL https://arxiv.org/abs/math/0507163.

R Tyrrell Rockafellar. *Convex analysis*, volume 28. Princeton university press, 1997.

C. Sakar and Yomi Kastro. Online Shoppers Purchasing Intention Dataset. UCI Machine Learning Repository, 2018. DOI: https://doi.org/10.24432/C5F88Q.

Aleksandra B. Slavković, Xiaotian Zhu, and Sonja Petrović. Fibers of multi-way contingency tables given conditionals: relation to marginals, cell bounds and markov bases, 2014. URL https://arxiv.org/abs/1401.1397.

Nils Strodthoff, Temesgen Mehari, Claudia Nagel, Philip J Aston, Ashish Sundar, Claus Graff, Jørgen K Kanters, Wilhelm Haverkamp, Olaf Dössel, Axel Loewe, et al. Ptb-xl+, a comprehensive electrocardiographic feature dataset. *Scientific data*, 10(1):279, 2023.

Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(1), August 2018. ISSN 2052-4463. doi: 10.1038/sdata.2018.161. URL http://dx.doi.org/10.1038/sdata.2018.161.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I Lunze, Wojciech Samek, and Tobias Schaeffter. Ptb-xl, a large publicly available electrocardiography dataset. *Scientific data*, 7(1):154, 2020.

Ian Waudby-Smith and Aaditya Ramdas. Confidence sequences for sampling without replacement, 2021. URL https://arxiv.org/abs/2006.04347.

Alan Yang, AmirEmad Ghassami, Maxim Raginsky, Negar Kiyavash, and Elyse Rosenbaum. Model-augmented estimation of conditional mutual information for feature selection, 2020. URL https://arxiv.org/abs/1911.04628.

Rico Zenklusen. Matching interdiction, 2008. URL https://arxiv.org/abs/0804.3583.

AI USE STATEMENT

Generative AI tools assisted with literature retrieval and synthesis; research ideation, methodology, and experimental design; software implementation and testing; analysis and interpretation of results; formulation of mathematical claims and development and writing of proofs; and drafting, editing, and polishing the manuscript. They also assisted in designing and implementing synthetic-data simulation generators. Synthetic samples were produced by executable code under the specified data-generating processes, not directly by an LLM. AI tools were not used to fabricate or alter observations in the original empirical datasets. Verification included automated software tests and symbolic and numerical consistency checks. Independent human review of the central proofs was completed before submission. The authors take responsibility for the final manuscript, mathematical claims, results, and artifacts, including all AI-assisted contributions.

GUIDE TO THE APPENDICES

Core reading route. For the two central claims, read the complete-joint Blackwell construction and fixed-rule characterization in Appendix A, the zero-link boundary in Appendix B, and the disclosure frontiers and joint bounds in Appendix C.2. These core proofs appear consecutively. Section S2 supplies the decision audits, including separate sufficiency and minimum-size checks for Heart; Section S5 identifies the inputs and replay scope. The observed-alignment appendix is the canonical ten-audit observed-alignment census.

Optional extension reading. Appendix D collects quantitative, smooth-forecast, weaker-release extensions, and empirical shattering witnesses. Later modules study stable refitting, online planning, sampling-based inference, and capacity sensitivity, followed by comparisons with neighboring information contracts. These modules use additional assumptions or different evaluation objects and are not prerequisites for the core frozen-rule certificates. The skin summary refers back to its protocol rather than repeating its census. All proofs and protocols remain available in the integrated appendix.

Artifact access. The accompanying anonymous code archive is `DeployRank_ICLR2027_submission_bundle.zip`; its `README.md` describes the sanitized contents. Section S5 distinguishes checks runnable from that archive from results needing original public data or restricted inputs. Author-side manifests and result ledgers retain full provenance but are not reviewer uploads. Restricted patient-level inputs and private contexts are not redistributed. The public-UCI Heart example includes derived losses and remapped context answers, not original clinical features or patient identifiers. The archive includes a self-contained Heart H–S replay package in `examples/heart_replay/`. Public-only verification checks the disclosed certificate without loading holder alignment; separate holder mode recomputes minimum-disclosure bounds. The archive's `evidence/claim_index.json` maps named theory audits to anonymous records; `evidence/witness_index.json` locates empirical witness summaries. `REPRODUCIBILITY.md` distinguishes direct replay, retained-outcome recount, and checks requiring additional inputs.

Four-claim lookup. The following are the canonical evidence pages for the shortest reading path.

Main question	Primary evidence	Page
Can one release admit every order?	Theorem 1's constructive Blackwell proof	14
When does a choice need no links?	Theorem 3's sharp upper-tail proof	18
What is the shortest holder proof?	Theorem 4's exact disclosure frontier	20
What changes on Heart H–S?	Table S10: choice, 86-link proof, 278-query discovery	65

Further entry points are Theorem 8 for local geometry, Appendix D.12 for the CIFAR witness, Table S13 for frozen cross-task transfer, and Appendix F.1 for public-type online planning. The CIFAR count/log-score aggregate replay is supplied in `examples/cifar_joint_replay/`; it does not reconstruct original image-level assignments or independently verify the Brier/spherical and TV summaries.

A Complete proof of deployment-rank shattering

Proof map and release contracts. The headline shattering result preserves the complete joint prediction law $P(Y, Z_{1:M})$ together with $P(Y, C)$; its constructive proof begins at "Proof of Theorem 1." The centered-tangent and finite-rule characterizations then identify when the same local geometry persists beyond the construction. The later Cartesian and raw-marginal results use the strictly weaker candidate-wise release $\{P(Y, C), P(Y, Z_1), \ldots, P(Y, Z_M)\}$ and are auxiliary: they establish useful identification boundaries but are not substitutes for the complete-joint theorem. The core route proceeds from definitions to Theorem 1 and the finite-rule characterization. Appendix D collects quantitative and weaker-release extensions and their executable audits.

Result	Extra premise and logical role	Loss dependence
Theorem 1	One complete-joint release and one common redundancy completion; strict contextual Blackwell orders.	Same worlds for all regular strictly proper losses.
Theorem 5	Quantitative value-gap refinement of a differentiable construction.	Release and tie may depend on the fixed loss.
Theorem 8	Full-rank centered derivative for a fixed smooth value map; the differentiable instance above supplies a witness.	Criterion applies to the chosen value map.
Corollary 5	Calibrated, overlapping smooth forecasts lifted from the fixed-score construction.	No shared cross-loss Blackwell lift is claimed.

In particular, the nonzero-derivative witness for Theorem 8 comes from the differentiable fixed-score instance, not by differentiating the redundancy minimum of Theorem 1.

A.1 Definitions

Let Y, C, and $Z_1, \ldots, Z_M$ have finite alphabets with at least two symbols, and fix one regular strictly proper scoring loss ℓ for predicting Y. Its Bayes risk

$$H_\ell(p) = \inf_a \ \mathbb{E}_{Y\sim p}[\ell(Y, a)]$$

is continuous and strictly concave on the outcome probability simplex. For a finite signal X define

$$R^*_\ell(X) = \sum_x P(x) H_\ell(P(Y = \cdot \mid X = x))$$

and $V^*_m = R^*_\ell(C) - R^*_\ell(C, Z_m)$. Concavity gives $V^*_m \geq 0$. Under strict propriety, equality holds exactly when the posterior given (C, Z_m) equals the posterior given C almost surely.

The headline release is $\widetilde{\mathcal{O}} = \{P(Y, C), P(Y, Z_{1:M})\}$, with compatible fiber $\mathcal{C}(\widetilde{\mathcal{O}})$. Write $\mathcal{O}_{\text{pair}} = \{P(Y, C), P(Y, Z_1), \ldots, P(Y, Z_M)\}$ for its candidate-wise marginalization; its intervals and margin are defined only with the auxiliary results below.

A.2 Proof of Theorem 1: score-uniform Blackwell shattering

Proof. The construction has five steps. Fix a full-support candidate-error law; vary only the conditional context-error law so that both released tables stay fixed; find one completion where every candidate is redundant given context; perturb posterior intervals into the requested nested Blackwell order; then control TV uniformly around that same completion. The requested permutation is chosen only after

the release and redundancy completion have been fixed. Fix $0 < r < 1/2$. Let Y be a fair bit, let $E_1, \dots, E_M$ be mutually independent Bernoulli bits independent of Y, with distinct rates

$$0 < r < q_1 < \cdots < q_M < 1/2, \qquad Z_m = Y \oplus E_m,$$

and define F and C by

$$P_\theta(F = 1 \mid E_{1:M} = e) = r + \sum_{m=1}^{M} \theta_m(e_m - q_m), \qquad C = Y \oplus F. \tag{4}$$

For each coordinate put

$$\theta_m^\star = \frac{r(1-r)(1-2q_m)}{q_m(1-q_m)(1-2r)}.$$

As $q_m \uparrow 1/2$, $\theta_m^\star \downarrow 0$. We may therefore choose the distinct q_m sufficiently close to $1/2$ that the right-hand side of (4) lies strictly in $(0,1)$ at $\theta = \theta^\star$, and throughout a fixed open box around it. Every atom of $(Y, E_{1:M}, F)$, hence of $(Y, C, Z_{1:M})$, then has positive probability.

Centering and independence give

$$P_\theta(F = 1) = r, \qquad P_\theta(F = 1, E_m = e) = P(E_m = e)\{r + \theta_m(e - q_m)\}.$$

Thus $P(Y, C)$ is the same BSC(r) law for every θ in the box, while $P(Y, Z_{1:M})$ is fixed before F is generated. The whole box therefore lies in one release fiber, and the pair (F, E_m) depends on θ only through θ_m.

Write $W_m = C \oplus Z_m = F \oplus E_m$ and $k_m = q_m(1 - q_m)$. Conditional on F, its binary channel is

$$P_\theta(W_m = 1 \mid F = 0) = q_m - \frac{\theta_m k_m}{1-r}, \qquad P_\theta(W_m = 1 \mid F = 1) = 1 - q_m - \frac{\theta_m k_m}{r}.$$

At $\theta_m = \theta_m^\star$ the two rows coincide, at

$$p_m^\star = \frac{q_m - r}{1 - 2r} \in (0, 1).$$

Hence $W_m \perp F$ at the single vector $\theta^\star$. Because Y is fair and independent of $(F, E_{1:M})$, the bit $C = Y \oplus F$ is independent of $(F, E_{1:M})$. Conditional on C, observing Z_m is equivalent to observing W_m, and $Y = C \oplus F$; therefore $Y \perp Z_m \mid C$ for every m at the same full-support completion $Q_0 = P_{\theta^\star}$. Every regular strictly proper score consequently assigns zero incremental value to every candidate there.

We next order the contextual experiments themselves, rather than one selected score. Put $\theta_m = \theta_m^\star - u_m$ with $u_m > 0$. Direct Bayes calculation gives the two posterior probabilities of F given W_m:

$$a_m(u_m) = r - \frac{u_m k_m (1-2r)}{1 - p_m^\star - 2u_m k_m}, \qquad b_m(u_m) = r + \frac{u_m k_m (1-2r)}{p_m^\star + 2u_m k_m}. \tag{5}$$

The denominators are positive after shrinking the open box. At $(q, u) = (1/2, 0)$, both coefficients of u in the deviations $r - a(q,u)$ and $b(q,u) - r$ equal $(1-2r)/2$. Continuity therefore permits the q-neighborhood and an upper bound on u to be chosen so that the largest of these deviation coefficients, over all admissible (q, u) and both sides, is strictly less than twice the smallest. It follows that whenever $u' \geq 2u > 0$,

$$[a(q', u'), b(q', u')] \supsetneq [a(q, u), b(q, u)]$$

for any two admissible rates q, q'.

For completeness, strict nesting yields an explicit outcome-independent garbling. Consider a more informative binary experiment A with output masses μ_0, μ_1 and posterior endpoints $a_0 < a_1$, and a nested experiment B with masses ν_0, ν_1 and $a_0 < b_0 < b_1 < a_1$. Define

$$K(1 \mid 0) = \frac{\nu_1(a_1 - b_1)}{\mu_0(a_1 - a_0)}, \qquad K(1 \mid 1) = \frac{\nu_1(b_1 - a_0)}{\mu_1(a_1 - a_0)},$$
$$K(0 \mid 0) = \frac{\nu_0(a_1 - b_0)}{\mu_0(a_1 - a_0)}, \qquad K(0 \mid 1) = \frac{\nu_0(b_0 - a_0)}{\mu_1(a_1 - a_0)}.$$

Bayes plausibility, $\mu_0 a_0 + \mu_1 a_1 = \nu_0 b_0 + \nu_1 b_1 = r$, shows that the two output probabilities sum to one for each input; strict nesting makes every entry lie in $(0, 1)$. The same identities show that applying K to A produces both the output masses and state–output joint probabilities of B. Thus B is a strict Blackwell garbling of A. In our construction, retain C, apply this kernel to W_m, and return $Z' = C \oplus W'$. The kernel never observes Y and maps the joint experiment (C, Z_m) exactly to the less informative one. Its strict interior mixing and the distinct source posteriors make strict Jensen's inequality apply to every strictly concave Bayes risk, without label symmetry.

Now fix a requested best-to-worst permutation π. For a scale $s > 0$, set

$$u_{\pi(j)} = s\, 2^{M-j}, \qquad j = 1, \dots, M.$$

Adjacent posterior intervals are strictly nested, so $(C, Z_{\pi(1)}) \succ_B \cdots \succ_B (C, Z_{\pi(M)})$. Given a radius η, take $s = 2\eta/(2^M - 1)$ and reduce the common radius η_0 until every u_m remains inside the probability box and below $\theta^\star_m$. Since the maps from (Y, E, F) to (Y, C, Z) are bijective for fixed Y,

$$\mathrm{TV}(P_{\theta^\star - u}, Q_0) = \mathbb{E}\left|\sum_m u_m(E_m - q_m)\right| \le \sum_m 2u_m q_m(1 - q_m) < \frac{1}{2}\sum_m u_m = \eta.$$

The release, Q_0, and η_0 were fixed before the permutation, radius, or loss was selected. Strict Blackwell dominance and strict concavity now give the requested value order simultaneously for every regular strictly proper loss. Finally, $q_1 < \cdots < q_M < 1/2$ makes $Z_1 \succ_B \cdots \succ_B Z_M$: each noisier BSC is a nontrivial garbling of the preceding one. This proves all claims. □

Exact construction audit. The released script `scripts/audit_score_uniform_blackwell.py` uses exact rational arithmetic, not evaluations of a finite score list. For $M \in \{2, 3, 5\}$ it checks every permutation at three radii: full support, exact preservation of both released laws, the common redundancy completion, strict nesting of every adjacent posterior interval, and every entry and identity of the explicit garbling kernels. All 384 permutation–radius cases pass; the machine-readable record accompanies the released audit code. This is a falsification audit of the algebra, not a replacement for the general proof.

The loss-specific quantitative refinement is stated and proved together in Appendix D.1 (Theorem 5); it is separate from the score-uniform Blackwell construction above.

A.3 Global-to-local characterization for finite fixed rules

Proof of Theorem 2. Set $T = \mathrm{lin}(\Gamma - \Gamma)$, $H = \mathbf{1}^\perp$, $\Pi = I - M^{-1}\mathbf{1}\mathbf{1}^\top$, $G = \Pi R$, and $K = G(\Gamma) = \mathcal{K}_\Gamma \subset H$. We prove equivalence of (i) every strict risk order being compatible, (ii) $0 \in \mathrm{int}_H K$, and (iii) a common tie $Q_0 \in \mathrm{ri}\,\Gamma$ with $\mathrm{rank}(\Pi D R|_T) = M - 1$. The set K is compact and convex. We first record a strict braid-chamber separation fact. For nonzero $a \in H$, choose a permutation π with $a_{\pi(1)} \ge \cdots \ge a_{\pi(M)}$. If $x_{\pi(1)} < \cdots < x_{\pi(M)}$, then $a^\top x < 0$. Indeed,

$$M a^\top x - (\mathbf{1}^\top a)(\mathbf{1}^\top x) = \sum_{i<j}(a_{\pi(i)} - a_{\pi(j)})(x_{\pi(i)} - x_{\pi(j)}) < 0.$$

Every summand is nonpositive and at least one is strictly negative because a is not constant.

Suppose (i) holds but $0 \notin \mathrm{int}_H K$. Finite-dimensional convex separation supplies nonzero $a \in H$, after changing its sign if needed, such that $a^\top x \ge 0$ for all $x \in K$. The chamber ordered oppositely to a lies strictly in $\{x : a^\top x < 0\}$ by the displayed fact, contradicting (i). Thus (i) implies (ii). Conversely, an H-open ball around zero meets every open cone, so (ii) implies (i).

Because G is affine, its linear part on the affine hull is $A = \Pi D R|_T : T \to H$. If (iii) holds, relative interior gives a ball $Q_0 + B_T(0, r) \subset \Gamma$. Surjectivity of A maps a sufficiently small T-ball onto an H-neighborhood of zero, proving (ii) and, moreover, placing every strict order arbitrarily near Q_0.

Finally assume (ii). Then K is full-dimensional in H, so $\mathrm{rank}\, A = M - 1$. Also $0 \in \mathrm{ri}\, K$. Relative interior commutes with linear images of convex sets (Rockafellar, 1997, Theorem 6.6); hence $G(\mathrm{ri}\,\Gamma) = \mathrm{ri}\, G(\Gamma)$ and some $Q_0 \in \mathrm{ri}\,\Gamma$ satisfies $G(Q_0) = 0$, equivalently $R(Q_0) = r_0\mathbf{1}$. This proves (iii). □

Implementation check. For an explicit rational polytope, write $\Gamma = \{q : Eq = e,\ Aq \le b\}$. An inequality belongs to the true affine hull exactly when its maximum feasible slack is zero. Append all such equalities to E, compute a basis N of the resulting tangent space, and test $\operatorname{rank}(\Pi D R N) = M - 1$. For the remaining inequalities I, solve

$$\max_{q,\delta} \delta \quad \text{s.t.} \quad Eq = e, \quad G(q) = 0, \quad A_I q + \delta \mathbf{1} \le b_I, \quad 0 \le \delta \le 1.$$

Its optimum is positive exactly when the tie lies in $\operatorname{ri}\Gamma$. The released checker separates numerical search from certification. Its floating-point LP/SVD path is only a screen and is never promoted to a theorem certificate. For decimal inputs, the validating path uses exact rational arithmetic: a positive certificate records an exact tie q_0 and $M - 1$ symmetric feasible direction pairs $q_0 \pm d_j$ whose centered risk images have rank $M - 1$. Their convex hull contains a neighborhood of zero in H, which proves the criterion without trusting a numerical affine-hull decision. Exact equality-rank deficiency can invalidate a positive numerical screen; otherwise an unverified screen is returned as unresolved. Chamber enumeration remains a small-instance falsification audit rather than an independent proof.

Corollary 3 (Exact universal ranking margin) *In the setting of Theorem 2, let $H = \mathbf{1}^{\perp}$, $\Pi = I - M^{-1}\mathbf{1}\mathbf{1}^{\top}$, and $K = \Pi R(\Gamma) = \mathcal{K}_\Gamma \subset H$. For each permutation π, define*

$$v^{\pi}_{\pi(k)} = k - \frac{M+1}{2}, \qquad k = 1, \dots, M, \qquad \mathcal{P}_M = \operatorname{conv}\{v^{\pi} : \pi \in S_M\}.$$

For every $g > 0$, every permutation has a compatible completion whose adjacent risk gaps are at least g if and only if $g\mathcal{P}_M \subseteq K$. Consequently, under the equivalent conditions of Theorem 2,

$$g_{\mathrm{all}} := \min_{\pi \in S_M} \max_{Q \in \Gamma} \min_{k < M} \{R_{\pi(k+1)}(Q) - R_{\pi(k)}(Q)\} = \max\{g \ge 0 : g\mathcal{P}_M \subseteq K\} > 0.$$

If an exact full-dimensional H-halfspace representation is $K = \{x \in H : a_s^{\top} x \le b_s,\ s = 1, \dots, S\}$, then

$$g_{\mathrm{all}} = \min_{s : \Pi a_s \neq 0} \frac{b_s}{h_{\mathcal{P}_M}(a_s)}, \qquad h_{\mathcal{P}_M}(a) = \frac{1}{2} \sum_{i<j} |a_i - a_j|.$$

A minimizing normal, sorted increasingly, identifies a worst-order certificate. For the release-level minimax regret in Equation (2)*,*

$$\mathfrak{G}(\Gamma) \ge (M-1) g_{\mathrm{all}}.$$

For any tied Q_0, any $0 \le \lambda \le 1$, and $\Gamma_\lambda = (1-\lambda)Q_0 + \lambda\Gamma$, $g_{\mathrm{all}}(\Gamma_\lambda) = \lambda g_{\mathrm{all}}(\Gamma)$.

Proof. Containment immediately supplies the completion with centered risk gv^{π}; centering leaves all risk differences unchanged. Conversely, select one centered witness $x^{\pi} \in K$ for each order, each with gaps at least g, and let $L = \operatorname{conv}\{x^{\pi} : \pi \in S_M\} \subseteq K$. Fix $a \in H$ and choose π such that $a_{\pi(1)} \le \cdots \le a_{\pi(M)}$. Then $z = x^{\pi} - gv^{\pi}$ is centered and ordered the same way, so

$$a^{\top} z = \frac{1}{M} \sum_{i<j} \{a_{\pi(j)} - a_{\pi(i)}\}\{z_{\pi(j)} - z_{\pi(i)}\} \ge 0.$$

The rearrangement inequality gives $a^{\top} v^{\pi} = h_{\mathcal{P}_M}(a)$, hence $h_L(a) \ge g h_{\mathcal{P}_M}(a)$. Support-function separation yields $g\mathcal{P}_M \subseteq L \subseteq K$. This also shows that a witness with exact gaps g exists for every order. In particular, for each fixed candidate a, choose an order placing a last and use its exact-gap witness at $g = g_{\mathrm{all}}$. Candidate a then has regret $(M-1)g_{\mathrm{all}}$, so taking the supremum over compatible completions and then the minimum over a proves the minimax-regret bound.

Classical permutahedron geometry gives $h_{\mathcal{P}_M}(a) = \frac{1}{2}\sum_{i<j} |a_i - a_j|$ (Postnikov, 2005); it is positive exactly when $\Pi a \neq 0$. Halfspace containment therefore gives the displayed ratio. If a_s attains the minimum and π sorts its coordinates, the same ordered-vector identity bounds that order's attainable minimum gap by $b_s / h_{\mathcal{P}_M}(a_s)$, proving the worst-order certificate. Finally, affinity and $G(Q_0) = 0$ give $G(\Gamma_\lambda) = \lambda G(\Gamma)$, proving exact scaling. □

The corollary concerns the convex completion fiber. It need not hold on the fixed-N lattice of exact empirical tables: for example, the centered image of $\Gamma = \Delta_3$ contains $(1/3)\mathcal{P}_3$, whereas empirical distributions with $N \le 2$ cannot realize a strict three-way order. Rational witnesses become exact rows after a suitable common-denominator enlargement; fixed-row realizability requires a separate integer certificate. Also, no fixed positive gap persists in every arbitrarily small neighborhood of a tie—the last display gives the exact linear decay.

Lemma 1 (Transfer from serialized coefficients) *Let r_m be the intended real-valued risk coefficient vector and $\widetilde{r}_m$ its serialized decimal representation on a probability simplex. If $\|r_m - \widetilde{r}_m\|_\infty \le \delta$ for every candidate, then for every feasible q,*

$$|\langle r_m, q\rangle - \langle \widetilde{r}_m, q\rangle| \le \delta, \qquad |(R_a - R_b) - (\widetilde{R}_a - \widetilde{R}_b)| \le 2\delta.$$

Consequently, a serialized strict gap larger than 2δ transfers with the same sign to the intended coefficients. An exact serialized tie transfers only to a 2δ approximate tie unless an independent analytic or interval calculation proves equality for the intended coefficients.

Proof. Because q is nonnegative and sums to one, $|\langle r_m - \widetilde{r}_m, q\rangle| \le \|r_m - \widetilde{r}_m\|_\infty$. Apply the triangle inequality to two candidates. □

Lemma 2 (Robust simplex transfer) *Let $\widetilde{v}_1, \dots, \widetilde{v}_L$ and $v_1, \dots, v_L$ lie in $H = \mathbf{1}^\perp$. If $rB_H \subseteq \operatorname{conv}\{\widetilde{v}_i\}$ and $\max_i \|v_i - \widetilde{v}_i\|_2 \le \eta < r$, then*

$$(r - \eta)B_H \subseteq \operatorname{conv}\{v_i\}.$$

If these vertices are centered affine risks of compatible completions in a convex fiber, their convex hull lies in that risk image. Consequently the fiber admits an intended-risk tie and local realization of every strict order in the continuous completion fiber. If uncentered M-candidate risk vertices differ coordinatewise by at most δ, one may take $\eta = \sqrt{M}\,\delta$ after centering.

Proof. For every unit $u \in H$, the support function of the perturbed hull obeys $h_v(u) \ge h_{\widetilde{v}}(u) - \eta \ge r - \eta$. The support-function characterization of compact convex containment gives the displayed inclusion. Centering is an orthogonal projection, so it cannot increase Euclidean norm, and $\|x\|_2 \le \sqrt{M}\|x\|_\infty$ gives the final statement. □

B SHARP DECISIONS FROM A FIXED UNLINKED RELEASE

This appendix proves Theorem 3 for a fixed unlinked release. Its decision cost is distinct from learning a sufficient release from two independent samples. Throughout this section, $C \in \{0, 1\}$, $X = (Y, Z_{1:M})$, and every fixed loss lies in $[0, B]$. The available release fixes P_X and $P_{Y,C}$ but not their row alignment. Write $p_y = P(C = 1 \mid Y = y)$ and, for an ordered candidate pair (a, j),

$$b_{aj}(X) = \ell_a(X, 0) - \ell_j(X, 0), \tag{6}$$
$$\delta_{aj}(X) = \ell_a(X, 1) - \ell_j(X, 1) - b_{aj}(X). \tag{7}$$

Every compatible completion is represented by a measurable $q(X) = P(C = 1 \mid X)$ satisfying $0 \le q \le 1$ and $E\{q(X) \mid Y = y\} = p_y$.

B.1 EXACT UPPER-TAIL BOUNDARY

For an integrable random variable V, define

$$\mathcal{T}_p(V) = \sup_{0 \le q \le 1,\ Eq = p} E(qV), \qquad 0 \le p \le 1.$$

Lemma 3 (Upper-tail primal–dual identity) *For every $p \in [0, 1]$,*

$$\mathcal{T}_p(V) = \inf_{\lambda \in \mathbb{R}} \{E(V - \lambda)_+ + p\lambda\}.$$

At an atom of V, fractional allocation on the boundary atom attains the primal. If V is an equal-weight vector of length n and $np = k$ is an integer, $\mathcal{T}_p(V)$ is the sum of its largest k entries divided by n.

Proof. For every feasible q and every λ, $E(qV) = E\{q(V - \lambda)\} + p\lambda \le E(V - \lambda)_+ + p\lambda$. For $p \in (0, 1)$, choose a $(1 - p)$ quantile λ and set q to one above that quantile, zero below it, and the required fraction on the boundary atom. This makes the inequality an equality. The endpoint cases are immediate. Sorting gives the finite-table claim. □

For each ordered pair, define its sharp worst compatible contrast

$$\Phi_{aj} = \sup_Q \{R_a(Q) - R_j(Q)\}.$$

Conditioning on Y and applying Lemma 3 gives

$$\Phi_{aj} = Eb_{aj} + \sum_y P(y)\mathcal{T}_{p_y}(\delta_{aj} \mid Y = y) \tag{8}$$

$$= \inf_{\lambda:\mathcal{Y}\to\mathbb{R}} E[b_{aj} + (\delta_{aj} - \lambda(Y))_+ + C\lambda(Y)]\,. \tag{9}$$

The first line is attained by aligning context one with the largest within-stratum increments. It is therefore an equality for the stated release fiber, not a relaxation. In an equal-weight empirical release it is an exact stratum-wise top-k_y calculation.

For a fixed candidate a, finite maxima commute with suprema:

$$\sup_Q \{R_a(Q) - \min_j R_j(Q)\} = \max_j \Phi_{aj}.$$

Consequently

$$\mathfrak{G} = \min_a \max_j \Phi_{aj}$$

is at most ϵ if and only if the release contains a candidate that is ϵ-optimal in every compatible completion. This proves Theorem 3.

The same identity separates a public contrast from the price of missing alignment:

$$\Phi_{aj} = \underbrace{Eb_{aj} + \sum_y P(y)p_y E(\delta_{aj} \mid y)}_{\text{conditional-independence contrast}} \tag{10}$$

$$+ \underbrace{\sum_y P(y)\left\{\mathcal{T}_{p_y}(\delta_{aj} \mid y) - p_y E(\delta_{aj} \mid y)\right\}}_{\text{alignment premium } \geq 0}. \tag{11}$$

Nonnegativity follows because the constant allocation $q = p_y$ is feasible. Thus a release certifies some candidate without row alignment at tolerance ϵ precisely when its conditional-independence contrast plus the alignment premium is at most ϵ against every challenger.

Beyond binary fixed contexts. For finite categorical C, the pairwise worst compatible contrast remains a transportation problem. Theorem 7 goes further: for a fixed candidate–challenger pair, its trimmed dual gives the exact minimum holder disclosure, is polynomial-time for every fixed alphabet size, and becomes NP-complete when the alphabet is part of the input. What is special to binary C is therefore the faster scalar upper-tail and top-k/prefix algorithm, not tractability itself for every fixed finite alphabet. Continuous context additionally requires measurable transport and a controlled approximation or regularity condition. Adaptive generators are covered only after their policy, sampling law, and evaluation object are frozen; unconstrained online generation is outside the fixed-rule guarantee.

C Deterministic disclosure certificates

The release and frozen loss table are fixed throughout this appendix. An exact certificate rules out every release-compatible alignment under which the returned candidate has regret greater than ϵ; it is not a sampling-confidence statement or a guarantee for a refitted rule.

Task	Scope and computational statement
Verify a fixed transcript	Explicit binary N-row table: stratum-wise top-k sums for each challenger; polynomial in the table size.
Minimum pairwise disclosure	One fixed candidate–challenger pair: binary prefix frontier in $O(N^2)$; fixed categorical alphabet in $N^{O(K)}\,\mathrm{poly}(L)$.
Minimum joint selection	One set must protect one candidate against all challengers; the binary count-grid is exact for shared within-stratum challenger orders and fixed stratum count. Without shared orders, threshold decision is NP-complete when candidate count is part of the input.
Optimal online policy	Adaptive worst-case discovery is a different problem; the public-type quotient is exact but its state count can be exponential.

C.1 Exact row-information cost

The disclosure results in this appendix and the online-discovery results in Appendix F concern a fixed finite release. "Exact" refers to optimization over every compatible row alignment under that release, not to a population-confidence statement.

Write $\rho_\epsilon^{>}(c^\star)$ for empirical TV distance to the strict finite-row bad set with regret $> \epsilon$. Proposition 7 instead uses $\rho_\epsilon(Q_0, a)$ for the closed continuous-law event with regret $\geq \epsilon$; the two radii can differ when an empirical lattice point attains the threshold exactly.

For a queried set S, the answers retain precisely $\mathcal{C}(S, c^\star) = \{c \in \mathcal{C} : c_S = c^\star_S\}$. Candidate a is exactly certified at tolerance ϵ if this restricted fiber contains no member of $\mathcal{B}_{a,\epsilon}$. Here the strict bad-alignment set and its row-disagreement set are

$$\mathcal{B}_{a,\epsilon} = \{c \in \mathcal{C} : R_a(c) - \min_j R_j(c) > \epsilon\}, \qquad D(c, c^\star) = \{i : c_i \neq c^\star_i\}.$$

Proposition 1 (Exact row-alignment certificate cost) *A queried set certifies candidate a exactly when it intersects the row-disagreement set of every member of $\mathcal{B}_{a,\epsilon}$. Hence its minimum size $\kappa_{a,\epsilon}(c^\star)$ is a hitting-set optimum and a transcript lower bound; the any-candidate cost is $\kappa_\epsilon^{\mathrm{select}}(c^\star) = \min_a \kappa_{a,\epsilon}(c^\star)$.*

Proof of Proposition 1. If S misses $D(c, c^\star)$ for some bad c, then $c_S = c^\star_S$, so that bad assignment remains in $\mathcal{C}(S, c^\star)$ and S cannot certify. If S intersects every bad-world disagreement set, each bad assignment conflicts with at least one revealed answer and none remains. This proves the hitting-set identity and its binary integer program. Relaxing $s_i \in \{0, 1\}$ to $0 \leq s_i \leq 1$ gives a lower bound; every integer feasible point gives a queried-set upper certificate.

For an adaptive zero-error strategy, fix its internal randomness and follow the transcript under $c^\star$. If the resulting queried set missed one bad-world disagreement set, that bad world would give the same answer at every query. Inductively the strategy would follow the same query path and emit the same certificate in that bad world, contradicting zero error. Thus its terminal queried set is a hitting set. Standard greedy set cover repeatedly queries a row hitting the most currently unexcluded explicit bad worlds and terminates with a feasible certificate. For a nonempty explicit bad-world universe, the standard set-cover analysis gives approximation factor $H_{|\mathcal{B}_{a,\epsilon}|} \leq 1 + \log |\mathcal{B}_{a,\epsilon}|$ relative to the exact certificate cost. This guarantee requires explicit access to the bad-world sets; it is not a guarantee for a release-visible heuristic on an implicit fiber. The executable audit reports both the greedy size and the exact integer optimum when enumeration is feasible.

If a strategy may return any candidate, apply the same argument to the candidate it returns on the truth's transcript. Its queried set has size at least that candidate's $\kappa_{a,\epsilon}$ and therefore at least $\kappa_\epsilon^{\mathrm{select}} = \min_a \kappa_{a,\epsilon}$. A candidate whose regret already exceeds ϵ at $c^\star$ has an infeasible certificate: the truth itself is a bad world with an empty disagreement set. □

C.2 Exact holder-side disclosure frontiers

Operational information contract and cost boundary. The holder knows the alignment; the evaluator knows the released losses and quotas. Authenticated selected links let the evaluator

recompute $\Phi_{aj}(S)$ without trusting an aggregate risk report. The cost counts explicit disclosure after alignment exists, not collection, linkage, authentication, or governance. This is the replayable-evidence contract of Section 4.2, not a preference over trusted aggregates.

C.2.1 Binary pairwise prefixes and joint count grids

Proof of Theorem 4. Fix a candidate–challenger pair and one outcome stratum. Any compatible completion differs from the true alignment by changing the same number q of true-zero rows to one and true-one rows to zero. If these sets are A and B, respectively, the change in the unnormalized pairwise contrast is

$$\sum_{i\in A}\Delta_i - \sum_{i\in B}\Delta_i.$$

After a disclosure, revealed rows cannot enter A or B. The adversary therefore pairs the largest remaining true-zero interactions with the smallest remaining true-one interactions and keeps precisely the profitable pairs.

Now condition on revealing exactly u true-zero and v true-one rows. If a revealed zero has smaller Δ than an unrevealed zero, exchanging their roles can only decrease every ordered profitable-pair sum. Repeating the exchange leaves the u largest true-zero interactions revealed. The symmetric argument reveals the v smallest true-one interactions. Thus the least possible remaining adversarial gain at (u, v) is exactly

$$E_y(u,v) = \sum_{r=1}^{\min(n_{y0}-u,\,n_{y1}-v)} [z_{y,u+r} - o_{y,v+r}]_+.$$

Outcome strata have separate fixed context counts, so their adversarial gains add. The true pairwise contrast plus $N^{-1}\sum_y E_y(u_y, v_y)$ is therefore the smallest sharp worst completion attainable with those stratum budgets. Minimizing the total budget subject to this quantity being at most ϵ proves the formula for $\kappa_{a,j}(\epsilon)$.

For computation, sorting costs $O(N \log N)$. Prefix sums and a monotone profitable-pair pointer evaluate all (u, v) cells in $O((n_{y0} + 1)(n_{y1} + 1))$ time per stratum. Min-plus convolution of the stratum frontiers costs $O(N^2)$ overall, as does the complete pairwise calculation. □

Proof of Corollary 2.

Any certificate that makes a safe against every challenger is in particular a certificate against each fixed j, hence has size at least $L_a = \max_{j\neq a} \kappa_{a,j}$. Conversely, the full union of one attaining certificate for every challenger is simultaneously safe. Because U_a^{union} is the smallest safe subcollection union,

$$U_a^{\mathrm{union}} \le \left|\bigcup_{j\neq a} S_{a,j}\right| \le \sum_{j\neq a}\kappa_{a,j} \le (M-1)L_a.$$

Minimizing over candidates that are ϵ-optimal at the truth gives the selection sandwich. If, for some a with $L_a = L$, one of the fixed size-L pairwise certificates is itself safe against all challengers, it is both a feasible selection certificate and a matching lower bound. It is also an admissible singleton in the fixed family defining U_a^{union}, proving $L = U_{\mathrm{union}} = \kappa_\epsilon^{\mathrm{select}}(c^\star)$.

It remains to prove the joint common-order clause. Fix a truth-feasible candidate a. Assume that for each outcome y there are row orders $z_{y,1}, \ldots, z_{y,n_{y0}}$ and $o_{y,1}, \ldots, o_{y,n_{y1}}$ satisfying, simultaneously for every $j \neq a$,

$$\Delta^{a,j}_{z_{y,1}} \ge \cdots \ge \Delta^{a,j}_{z_{y,n_{y0}}}, \qquad \Delta^{a,j}_{o_{y,1}} \le \cdots \le \Delta^{a,j}_{o_{y,n_{y1}}}.$$

For any disclosure S with u_y revealed true zeros and v_y revealed true ones in stratum y, apply the preceding exchange separately to each revealed row. Because the same weak order is valid for every challenger, replacing S by the corresponding shared prefixes weakly decreases every challenger's sharp worst contrast at once, without changing the disclosure size. Hence an optimal unrestricted multi-challenger certificate has shared-prefix form. For such a prefix define

$$E_{y,j}(u,v) = \sum_{r=1}^{\min(n_{y0}-u,\,n_{y1}-v)} \left[\Delta^{a,j}_{z_{y,u+r}} - \Delta^{a,j}_{o_{y,v+r}}\right]_+.$$

Its exact unnormalized worst contrast is $C^*_{a,j} + \sum_y E_{y,j}(u_y, v_y)$, where $C^*_{a,j} = N[R_a(c^\star) - R_j(c^\star)]$. Thus minimizing the total prefix budget subject to $C^*_{a,j} + \sum_y E_{y,j}(u_y, v_y) \le N\epsilon$ for every challenger gives the exact joint optimum in Corollary 2, over arbitrary disclosed sets rather than only over prefix sets. Equivalently,

$$\kappa_{a,\epsilon} = \min_{\mathbf{u},\mathbf{v}} \left\{ \sum_y (u_y + v_y) : C^*_{a,j} + \sum_y E_{y,j}(u_y, v_y) \le N\epsilon \text{ for every } j \ne a \right\}.$$

The diagonal recurrence

$$E_{y,j}(u, v) = [\Delta^{a,j}_{z_{y,u+1}} - \Delta^{a,j}_{o_{y,v+1}}]_+ + E_{y,j}(u + 1, v + 1)$$

with zero boundary values fills all local grids in $O(MN^2)$ time. Here H counts outcome strata, not the centered-risk hyperplane in Appendix A. There are at most $(N + 1)^{2H}$ joint count vectors, each checked against at most $M - 1$ challengers, giving $O(MN^{2H})$ time for fixed H. To obtain the sharper factor, choose one pairwise-optimal prefix count vector for each challenger and take their coordinatewise maximum. The resulting shared prefix contains every pairwise prefix and is jointly sufficient. Its size is at most both $\sum_{j \ne a} \kappa_{a,j} \le (M - 1)L_a$ and the sum of $2H$ coordinatewise maxima, each at most L_a. Thus $\kappa_{a,\epsilon} \le \min\{M - 1, 2H\}L_a$.

Finally, if a challenger b has common-prefix worst contrast at least that of every challenger for every count vector, every prefix safe against b is jointly safe. A pairwise-optimal prefix for b therefore attains the joint lower bound, so $\kappa_{a,\epsilon} = \kappa_{a,b} = L_a$. Weak orders handle ties without arbitrary challenger-specific tie breaking. □

C.2.2 Unrestricted joint selection and hardness

Why joint optimization is not merely a union. One outcome stratum contains true-zero rows z_1, z_2, z_3 and true-one rows o_1, o_2, o_3. At tolerance zero, let candidate a have loss $1/2$ and let two challengers have losses

$$\ell_b(i, c) = \frac{13}{24} - \frac{c}{2}\mathbf{1}\{i \in \{z_1, z_2\}\}, \qquad \ell_d(i, c) = \frac{3}{8} + \frac{c}{2}\mathbf{1}\{i \in \{o_1, o_2\}\}.$$

Both true unnormalized contrasts for a are $-1/4$, and the displayed row orders are common extremal orders. The unique minimum certificate against b is $\{z_1, z_2\}$; against d it is $\{o_1, o_2\}$. Thus independently optimal pairwise certificates have union size four. No two-row set certifies against both challengers, whereas revealing all three true-zero rows fixes the alignment through the released count. Consequently

$$\kappa_{a,b} = \kappa_{a,d} = 2, \qquad \kappa_{a,0} = 3, \qquad |S_{a,b} \cup S_{a,d}| = 4.$$

All losses lie in $[0, 1]$, and neither pairwise optimum has a tie. This example shows that the exact joint count-grid is strictly sharper than unioning pairwise optima, while joint comonotonicity alone need not make the pairwise lower bound exact.

This is a holder-side result: construction uses the true context of candidate rows, while verification of the disclosed certificate uses only the public release and disclosed answers. The reference implementation is `src/deployrank/pairwise_disclosure_frontier.py`. A separate script that imports no DeployRank module recomputes the top-k completions, verifies the returned row sets, and matches exhaustive subset search on 1,000 random finite problems.

Proposition 2 (Exact unrestricted joint-disclosure program) *Fix a truth-feasible candidate a and binary context. For challenger j, let $t^j_i = d^{a,j}_i(1) - d^{a,j}_i(0)$ and $T_j = \sum_i d^{a,j}_i(c^\star_i)$. For outcome stratum h, put $m_{jh} = \min_{i:y_i=h} t^j_i$ and $M_{jh} = \max_{i:y_i=h} t^j_i$. The unrestricted minimum joint disclosure is the optimum of the following mixed-integer linear program. Let $x_i \in \{0, 1\}$ indicate disclosure, $\theta_{jh} \in [m_{jh}, M_{jh}]$, and $q_{ij} \ge 0$. Minimize $\sum_i x_i$ subject, for every challenger $j \ne a$, to*

$$T_j + \sum_i q_{ij} \le N\epsilon,$$

and, for each row i in stratum h,

$$q_{ij} \ge t^j_i - \theta_{jh} - (t^j_i - m_{jh})x_i, \quad c^\star_i = 0,$$
$$q_{ij} \ge \theta_{jh} - t^j_i - (M_{jh} - t^j_i)x_i, \quad c^\star_i = 1.$$

It has N binary variables, $(M-1)(N+H)$ continuous variables, and $O(MN)$ inequalities. Exactness of the formulation does not imply polynomial-time solvability.

Proof. For a disclosed set $S = \{i : x_i = 1\}$, let Z_h and O_h be its undisclosed true-zero and true-one rows in stratum h. The sharp adversarial residual for challenger j is

$$\sum_h \min_\theta \left\{ \sum_{i \in Z_h} [t_i^j - \theta]_+ + \sum_{i \in O_h} [\theta - t_i^j]_+ \right\}.$$

Indeed, a quota-preserving alternative matches equally many true-zero rows moved to one with true-one rows moved to zero; sorting the former downward and the latter upward gives the sum of positive paired differences. A threshold between the last profitable pair and the first nonprofitable pair gives exactly the displayed hinge minimum, including ties and empty groups. When $x_i = 0$, the linear inequalities are the two hinge epigraphs. When $x_i = 1$, their big-M terms make the row contribution zero for every allowed θ_{jh}. Minimizing $\sum_i x_i$ while bounding each exact worst contrast by $N\epsilon$ therefore ranges over all disclosed row sets and is exact. □

Theorem 6 (Complexity boundary for unrestricted joint disclosure) *Suppose losses are explicitly given rational numbers and a disclosed row reveals its true binary context. With equally weighted rows, one outcome stratum, a fixed context-one count, and $\epsilon = 0$, each of the following decision problems is NP-complete when the number of candidates is part of the input: whether at most k disclosed rows make a specified candidate optimal under every compatible alignment, and whether at most k disclosed rows make some candidate optimal under every compatible alignment. This remains true with all losses in $[0, 1]$ and a unique true winner. Consequently, computing either minimum disclosure size is NP-hard.*

Proof. For membership in NP, a certificate gives a disclosed set S and, for the selection problem, a candidate a. Let U be the undisclosed rows and let $q = K - \sum_{i \in S} c_i^\star$ be their remaining context-one quota. For each challenger j, put

$$\delta_i(c) = \ell_a(i, c) - \ell_j(i, c), \qquad h_i = \delta_i(1) - \delta_i(0).$$

If $h_{(1)} \geq \cdots \geq h_{(|U|)}$ are the sorted undisclosed interactions, the sharp unnormalized worst contrast is

$$\sum_{i \in S} \delta_i(c_i^\star) + \sum_{i \in U} \delta_i(0) + \sum_{r=1}^{q} h_{(r)}.$$

Thus universal optimality is checked for every challenger by polynomial-time sorting and exact rational comparison.

For hardness, reduce from VERTEX COVER (Karp, 1972). Let $G = (V, E)$ be a simple graph with at least one edge and set

$$N = |V| + 2|E|, \qquad K = 2|E|, \qquad s = \tfrac{1}{4}, \qquad b = -s/N.$$

Create one true-zero row z_v for every vertex and two true-one rows $o_{e,1}, o_{e,2}$ for every edge. Candidate a has constant loss $1/2$. For every edge $e = \{u, v\}$ create a challenger j_e with

$$\ell_{j_e}(i, c) = \tfrac{1}{2} - b - t_i^e c, \qquad t_i^e = \begin{cases} s, & i \in \{z_u, z_v\}, \\ -s, & i \in \{o_{e,1}, o_{e,2}\}, \\ 0, & \text{otherwise.} \end{cases}$$

The losses share denominator $4N$ and lie in $[0, 1]$: their range is contained in $[1/4 + 1/(4N),\ 3/4 + 1/(4N)]$, and $N \geq 4$.

Write $F_e = \{z_u, z_v, o_{e,1}, o_{e,2}\}$ and let $r_e(c)$ count rows in F_e whose context differs from truth. The unnormalized contrast is

$$D_e(c) = \sum_i [\ell_a(i, c_i) - \ell_{j_e}(i, c_i)] = s\{r_e(c) - 3\}.$$

At truth, $D_e = -3s$ for every edge, so a is the unique winner. Challenger j_e strictly defeats a exactly when all four rows of F_e change; three changes yield a tie, which is safe for ordinary $\epsilon = 0$ optimality.

If S intersects every F_e, then $r_e(c) \le 3$ for every compatible completion, so a remains optimal. Conversely, if $S \cap F_e = \varnothing$, flip precisely the two zero and two one rows in F_e. This preserves the quota and the disclosure, and gives $D_e = s > 0$. Hence S certifies a iff it hits every F_e.

A vertex cover C gives the hitting disclosure $\{z_v : v \in C\}$. Conversely, retain the vertex rows in any hitting disclosure and replace each selected auxiliary row $o_{e,r}$ by either endpoint of its edge, removing duplicates. An auxiliary row occurs only in its own edge set, so the resulting vertices cover G and are no more numerous. Therefore the fixed-candidate minimum is exactly the vertex-cover number $\tau(G)$.

Finally, truth remains compatible with every truthful disclosure. Since a is uniquely optimal there, no challenger is optimal throughout the remaining fiber; the selection and fixed-candidate minima both equal $\tau(G)$. The same threshold k proves both reductions. □

The exact-integer audit enumerates every quota-compatible alignment, every disclosure, and every candidate for K_2, K_2 plus an isolate, P_3, K_3, and two disjoint edges. In all five cases the two disclosure minima equal the vertex-cover number. The author-side machine-readable record is `results/joint_disclosure_complexity_v1/exact_finite_audit.json`.

C.2.3 FINITE-CATEGORICAL PAIRWISE FRONTIER

Theorem 7 (Finite-categorical disclosure frontier) *Assume equal-weight identified rows, a finite context alphabet $[K]$, exact released outcome–context counts, complete finite rational loss tables, and a holder-known true alignment $c^\star$. For a fixed candidate–challenger pair put $d_i(c) = \ell_a(i,c) - \ell_j(i,c)$. Within every nonempty outcome stratum y, define*

$$D_y = \max_{i:y_i=y} \max_{c,c' \in [K]} |d_i(c) - d_i(c')|, \quad \mathcal{V}_y = \{v \in \mathbb{R}^K : v_K = 0, |v_c - v_{c'}| \le D_y\},$$

and

$$g_i(v) = \max_{c \in [K]} \{d_i(c) - d_i(c_i^\star) + v_{c_i^\star} - v_c\}.$$

If $g_{y,(1)}(v) \le \cdots \le g_{y,(n_y)}(v)$ are the ordered row gains, the exact minimum residual adversarial gain after disclosing k truthful rows in stratum y is

$$E_y(k) = \min_{v \in \mathcal{V}_y} \sum_{r=1}^{n_y - k} g_{y,(r)}(v).$$

Consequently,

$$\kappa_{a,j}(\epsilon) = \min_{\{k_y\}} \left\{ \sum_y k_y : R_a(c^\star) - R_j(c^\star) + N^{-1} \sum_y E_y(k_y) \le \epsilon \right\}.$$

Set $E_y(0) = 0$ for an empty stratum. For every fixed K, all frontiers and attaining disclosure sets are computable in $N^{O(K)}\,\mathrm{poly}(L)$ time, where L is the rational-input encoding length. When K is part of the input, deciding whether $\kappa_{a,j}(0) \le k$ is NP-complete, even with one outcome stratum, two candidates, one row per context category, rational losses in $[0,1]$, and a unique true winner. The any-candidate selection version is NP-complete under the same restrictions.

Proof. Fix one stratum and a disclosed set S, and write U for the undisclosed rows. Its residual count of category c is $r_c = |\{i \in U : c_i^\star = c\}|$. The adversary's assignment problem is an integral transportation problem. Its linear-program dual, after eliminating one row potential per $i \in U$, is

$$\min_{v \in \mathbb{R}^K} \left\{ \sum_{i \in U} \max_c [d_i(c) - v_c] + \sum_c r_c v_c \right\}.$$

The transportation polytope is integral, so this equals the maximum over finite row assignments, not merely a fractional relaxation. Because $\sum_c r_c v_c = \sum_{i \in U} v_{c_i^\star}$, subtracting the true undisclosed contrast gives exactly $\min_v \sum_{i \in U} g_i(v)$.

It remains to justify the compact potential domain, including categories with zero residual count. Take an optimal transport dual (u, v) satisfying $u_i + v_c \ge d_i(c)$ and replace

$$v_c \quad \text{by} \quad v_c' = \max_{i \in U} \{d_i(c) - u_i\}.$$

Then $v' \le v$, dual feasibility is preserved, and the objective cannot increase because every $r_c \ge 0$; optimality is therefore preserved. The elementary inequality $|\max_i a_i - \max_i b_i| \le \max_i |a_i - b_i|$ gives $|v'_c - v'_{c'}| \le D_y$. Translation of all potentials leaves the objective unchanged, so normalize $v'_K = 0$. If U is empty the residual gain is zero directly. Thus an optimum always lies in $\mathcal{V}_y$.

For a fixed budget k, minimization over the disclosed set and the dual potential commutes:

$$\min_{|S|=k} \min_{v \in \mathcal{V}_y} \sum_{i \notin S} g_i(v) = \min_{v \in \mathcal{V}_y} \min_{|S|=k} \sum_{i \notin S} g_i(v).$$

At fixed v, the inner minimum discloses the k largest row gains, leaving the sum of the $n_y - k$ smallest. This proves the displayed $E_y(k)$; separate strata add, and min-plus convolution over their budgets proves the formula for $\kappa_{a,j}(\epsilon)$.

For fixed K, each g_i is the maximum of K rational affine functions on the compact polytope $\mathcal{V}_y$, whose dimension is at most $K - 1$. If $D_y = 0$, this polytope is the singleton $\{0\}$ and is evaluated directly. Otherwise form the arrangement of all equality hyperplanes between the $n_y K$ affine pieces together with the facets of $\mathcal{V}_y$. Inside each cell, every maximizing branch and the ordering of the g_i are fixed, so every budget objective is affine and attains its minimum at a vertex of the closure of the bounded subdivision. After discarding equality identities and infeasible constant equalities, enumerating intersections of at most $K - 1$ independent hyperplanes, rejecting infeasible points, and evaluating all ordered gains computes every $E_y(k)$ exactly. There are $O(n_y^2 K^2 + K^2)$ hyperplanes, giving $N^{O(K)}$ arithmetic time; rational linear systems have polynomial bit complexity in L for fixed K. Cross-stratum min-plus convolution adds $O(N^2)$ time. This is an XP result in K, not an FPT claim; the binary $O(N^2)$ prefix algorithm remains much faster.

For hardness, reduce from VERTEX COVER. Given a graph $G = (V, E)$ with $N = |V| = K$, create one row and one context category for each vertex, use one outcome stratum, quota one per category, and truth $c_i^\star = i$. Let $s = 1/4$, $b = -s/N$, and

$$d_i(c) = \begin{cases} b, & c = i, \\ b + s, & \{i, c\} \in E, \\ b - s, & c \ne i \text{ and } \{i, c\} \notin E. \end{cases}$$

Set $\ell_a(i, c) = 1/2$ and $\ell_j(i, c) = 1/2 - d_i(c)$; all losses lie in $[0, 1]$. At truth the total contrast is $Nb = -s$, so a is the unique winner. If a disclosure S misses edge $\{i, j\}$, swapping the two undisclosed endpoint categories preserves every quota and changes the total contrast to $+s$, so certification fails. If S is a vertex cover, its undisclosed vertices form an independent set. Every nonidentity compatible permutation then moves rows only along nonedges, decreasing the total contrast by s per moved row, so a remains optimal. Hence the minimum disclosure is exactly the vertex-cover number.

Membership in NP follows because a proposed disclosure is verified by one rational transportation problem in polynomial time. Truth remains compatible after every truthful disclosure and uniquely favors a, so no challenger can be universally optimal; the any-candidate version has the same minimum. This proves both NP-completeness statements. □

Binary specialization. For $K = 2$, label the contexts $0, 1$, write $\Delta_i = d_i(1) - d_i(0)$, and set $t = v_1 - v_0$. Then the categorical row gain is

$$g_i(t) = [\Delta_i - t]_+ \quad (c_i^\star = 0), \qquad g_i(t) = [t - \Delta_i]_+ \quad (c_i^\star = 1).$$

These scalar hinge residuals are exactly the profitable-pair geometry used by Theorem 4. Thus the categorical theorem extends the object being optimized, while the binary sorting theorem supplies the sharper $O(N^2)$ algorithm.

The numerical implementation uses a compact holder-side MILP and independently rechecks its returned set by a transportation assignment; solver optimality is not presented as an exact-arithmetic proof, and any verified value inside the declared threshold band returns NUMERICALLY UNRESOLVED. A deterministic audit computes all 356 budget entries on 60 random $K = 2, 3, 4$ problems, checks transport primal–dual equality for 3,160 disclosed subsets, and matches exact vertex-cover values on five reduction instances. Three additional full-frontier cases cover zero quotas, exhausted categories, ties, constant losses, multiple outcomes, and exact threshold equality.

The code and author-side record are `src/deployrank/categorical_disclosure.py` and `results/categorical_disclosure_frontier_v1/exact_small_audit.json`.

At realistic scale, a post-hoc 4,000-row COIL stress test with $C = \min(\texttt{APERSAUT}, 2)$ gives $71 \le \kappa^{\mathrm{select}}_{.005} \le 238$. The global lower endpoint minimizes all-candidate pairwise floors; the public upper certificate discloses all 152 category-2 rows plus 86 residual binary-fiber rows, and exact dyadic replay gives worst contrasts 0.004723379, 0.004853934, and 0.003860540. The public artifact contains no undisclosed contexts. Because the construction exhausts the small third category, it is evidence for categorical sufficiency at large N, not for practical scaling of the general $N^{O(K)}$ algorithm.

C.2.4 Online-query contraction floor

A release-only contraction bound for online transcripts. The exact frontier above answers minimum disclosure after the holder knows the alignment. For a policy that does not know it, the following pathwise contraction bound remains useful.

Proposition 3 (Release-visible query-contraction floor) *Assume binary context, equal row weights, fixed outcome–context counts, and fixed losses. Put $d_i^{a,j}(c) = L_{i,a}(c) - L_{i,j}(c)$ and $\Delta_i^{a,j} = d_i^{a,j}(1) - d_i^{a,j}(0)$. Define*

$$\beta_{a,i} = \frac{1}{n} \max_j \max_{h:y_h=y_i} |\Delta_i^{a,j} - \Delta_h^{a,j}|, \qquad \beta_{a,j,i} = \frac{1}{n} \max_{h:y_h=y_i} |\Delta_i^{a,j} - \Delta_h^{a,j}|.$$

Let $W_a(S)$ be a's sharp worst-case regret and $W_{a,j}(S)$ the sharp maximum of $R_a - R_j$ after transcript S. Parenthesized subscripts denote decreasing order. Every transcript that certifies a at tolerance ϵ satisfies simultaneously

$$\sum_{r\le|S|} \beta_{a,(r)} \ge [W_a(\varnothing) - \epsilon]_+, \qquad \sum_{r\le|S|} \beta_{a,j,(r)} \ge [W_{a,j}(\varnothing) - \epsilon]_+ \quad (\forall j).$$

Let B_a have rows $(\beta_{a,i})_i$ and $(\beta_{a,j,i})_i$, and let b_a collect the corresponding right-hand sides above. Every zero-error any-candidate policy has stopping depth at least

$$q_\epsilon^{\mathrm{ctr}} = \min_a \min_{x\in\{0,1\}^n} \{\mathbf{1}^\top x : B_a x \ge b_a\}.$$

This release-only covering MILP is neither a policy nor an online optimum. It can strictly dominate separate sorting; on Waterbirds one challenger binds, so the joint step does not further raise the challenger-specific floors.

Proof. Consider one feasible reveal $i \mapsto c$ from a current transcript. Fix a challenger and a completion maximizing its pairwise loss difference before the reveal. If that completion already assigns $c_i = c$, it remains feasible. If not, fixed context counts and nonempty child feasibility imply an unqueried h in the same outcome stratum with the opposite assignment. Swapping the two contexts produces a child-feasible completion and changes the normalized pairwise objective by at most $|\Delta_i^{a,j} - \Delta_h^{a,j}|/n \le \beta_{a,i}$. Taking the maximizing challenger gives $W_a(S\cup\{i\}) \ge W_a(S) - \beta_{a,i}$. Telescoping over the transcript and upper-bounding its contractions by the $|S|$ largest $\beta_{a,i}$ proves the first display. Holding j fixed throughout the same exchange argument proves the second display with $\beta_{a,j,i}$. The indicator of the terminal queried set is therefore feasible for the joint covering problem of the returned candidate. Minimizing first over feasible indicators and then over candidates proves the floor. Enforcing all constraints jointly contains every separately sorted lower-bound relaxation, proving dominance. □

Proof of Corollary 1. For a candidate a, let $W_a(S)$ be its sharp worst compatible regret after transcript S. Initially $W_a(\varnothing) \ge G_0$ for every a. In the binary-context setting of Proposition 3, each pairwise loss difference lies in $[-B, B]$, so each interaction contrast $\Delta_i^{a,j}$ lies in $[-2B, 2B]$. Therefore $\beta_{a,i} \le 4B/N$ for every row and candidate. The proposition's single-reveal exchange argument is pathwise, and telescoping along an arbitrary adaptive transcript of length q gives

$$W_a(S) \ge W_a(\varnothing) - \frac{4Bq}{N} \ge G_0 - \frac{4Bq}{N}.$$

If the policy stops and certifies its returned candidate a at tolerance ϵ, then $W_a(S) \le \epsilon$. Rearrangement and integrality of q give the claimed ceiling. The argument applies to every terminal path and therefore also to the worst-case depth of any zero-error adaptive policy. □

Corollary 4 (Radius–certificate separation and fixed-threshold cost) *For every $n = 2k \geq 4$ and $0 < \epsilon_n < 1/(2n)$, a bounded-loss binary-context release has nearest strict bad-world TV $2/n$ but exact certificate cost $n/2$. For fixed $0 < \epsilon < 1/6$, a variant has a nonvanishing bad-world radius and linear truth-known certificate cost. For even n with $n\epsilon > 1$, put $U = \lfloor 3n\epsilon \rfloor$ and $L = \lceil n\epsilon \rceil$. Its exact adaptive zero-error minimax discovery cost and worst truth-known certificate complexity are*

$$D^*_{\text{online}} = k - U + L - 1, \qquad \max_{c^\star} \kappa^{\text{select}}_\epsilon(c^\star) = \max\{k - U, L\}.$$

Their gap is $\min\{k - U, L\} - 1$, positive and linear in n for fixed $\epsilon \in (0, 1/6)$ and all sufficiently large even n. On the displayed truth, the certificate is $k - U$ and the gap is exactly $L - 1$.

Proof of Corollary 4. Partition $n = 2k$ identified rows into A and B, each of size k. Let the release fix one outcome, a distinct public prediction profile for every row, and exactly k context-one rows. The true assignment has context one on A and zero on B. Candidate a has row loss $L_a(i, c) = \mathbf{1}\{i \in B, c = 1\}$; candidate b has constant row loss $1/(2n)$. Both losses lie in $[0, 1]$. At the truth, $R_a = 0$ and $R_b = 1/(2n)$, so a wins.

Every other compatible assignment swaps $t \geq 1$ rows from A with t rows from B. Then $R_a = t/n$, so its regret is $t/n - 1/(2n)$. For any $0 < \epsilon_n < 1/(2n)$ every such world is bad for selecting a. A one-pair swap is nearest: it changes two row contexts, which is TV $2/n$ on the empirical law over identified rows and context. The minimal bad-world disagreement sets are all pairs $\{i, j\}$ with $i \in A$ and $j \in B$. Hitting every such pair is a vertex cover of the complete bipartite graph $K_{k,k}$; querying all of either side is sufficient, while fewer than k queried rows leave one unqueried endpoint on both sides and hence an unhit edge. Therefore $\kappa_{a,\epsilon_n}(c^\star) = k = n/2$. The row bipartition is public. Hence a valid nonadaptive online policy can query all rows in A (or all rows in B) and then run the exact fail-closed audit. On the displayed truth those k answers hit every dangerous disagreement set and certify a; on any other transcript the audit either certifies correctly or continues. Thus the adaptive transcript lower bound is attained on $c^\star$.

For the fixed-threshold variant, fix $0 < \epsilon < 1/6$ and instead give candidate b constant loss 2ϵ. Candidate a is uniquely ϵ-optimal at the truth, because b has regret 2ϵ. A swap world is bad for selecting a exactly when $t/n - 2\epsilon > \epsilon$, so the smallest bad swap size is $t_{n,\epsilon} = \lfloor 3n\epsilon \rfloor + 1$. It changes $2t_{n,\epsilon}$ row contexts and hence lies at empirical TV $2t_{n,\epsilon}/n$. A queried set misses a minimal dangerous world exactly when its complement contains at least $t_{n,\epsilon}$ rows of both A and B. Therefore it hits every dangerous disagreement set exactly when fewer than $t_{n,\epsilon}$ unqueried rows remain on at least one side. The minimum is attained by querying $k - t_{n,\epsilon} + 1$ rows of one side, and no smaller set suffices. Thus

$$\kappa^{\text{select}}_\epsilon(c^\star) = \kappa_{a,\epsilon}(c^\star) = k - t_{n,\epsilon} + 1, \qquad \rho^{>}_\epsilon(c^\star) = \frac{2t_{n,\epsilon}}{n}.$$

Candidate b has no valid certificate on this truth, so taking the minimum over candidates cannot bypass the lower bound. Dividing by n and taking even $n \to \infty$ gives the stated limits.

For the online statement, write r for the number of context-one rows in B, $U = \lfloor 3n\epsilon \rfloor$, and $L = \lceil n\epsilon \rceil$. Candidate a is ϵ-optimal exactly when $r \leq U$, while b is ϵ-optimal exactly when $r \geq L$. The two integer decision regions cover all worlds because $L \leq U + 1$. Querying rows of B in any fixed public order and stopping for b after L ones or for a after $k - U$ zeros is zero-error. Its worst-case depth is $(L - 1) + (k - U - 1) + 1 = k - U + L - 1$.

It remains to rule out a better policy that mixes the two public sides. After an arbitrary transcript, let a_1, a_0, b_1, b_0 count observed ones and zeros in A and B. The compatible values of r form the integer interval

$$[\max(b_1, a_0),\ k - \max(a_1, b_0)].$$

Hence a is certified exactly when $\max(a_1, b_0) \geq k - U$, and b exactly when $\max(b_1, a_0) \geq L$. Set $p = k - U - 1$, $q = L - 1$, and $D = p + q + 1 = k - U + L - 1$. Since $L \leq U + 1$, $D \leq k$. At any feasible nonterminal transcript of depth below $D - 1$, either binary answer to the next query is globally completable to exactly k ones: the queried side is not exhausted and fewer than k total ones and zeros have appeared. If both answers stopped after a query in A, the current transcript would already have $a_1 = p$ and $a_0 = q$, hence at least $p + q = D - 1$ queries; the same argument for B uses $b_0 = p$ and $b_1 = q$. Thus at least one answer remains feasible and nonterminal. An adversary can preserve ambiguity through $D - 1$ answers, so every zero-error policy requires a Dth query in

the worst case. This matches the one-side policy. On the displayed truth the policy observes only zeros in B and stops after $k - U = k - t_{n,\epsilon} + 1$, equal to the post-hoc certificate cost.

The same terminal conditions give the truth-known cost for every compatible truth. If $r < L$, only a is ϵ-optimal and a certificate needs exactly $k - U$ observed zeros in B (equivalently, ones in A). If $r > U$, only b is ϵ-optimal and exactly L observed ones in B (equivalently, zeros in A) are necessary and sufficient. In the overlap $L \le r \le U$, either candidate may be certified, so the minimum is $\min\{k - U, L\}$. The extreme truths $r = 0$ and $r = k$ attain the first two costs, proving $\max_{c^\star} \kappa_\epsilon^{\mathrm{select}}(c^\star) = \max\{k - U, L\}$ and the displayed minimax–certificate gap.

□

The threshold condition is essential only for simultaneously obtaining vanishing TV radius and $n/2$ cost. At fixed positive ϵ, bounded-loss continuity forces the nearest-danger radius to stay nonzero. The variant above still requires a linear number of exact row queries to return any qualified candidate whenever $\epsilon < 1/6$.

In general, certificate size is an oracle property of the realized alignment, not an online policy cost. Theorem 12 nevertheless characterizes the exact minimax online policy for every finite binary fixed-rule release by a Bellman recursion over public row-type counts. Its state bound can be prohibitive for native continuous losses; the fixed-margin family above is the structured case where that online optimum has a closed form.

D Extensions and release-specific empirical witnesses

D.1 Proof of Theorem 5: quantitative fixed-score refinement

Write $v_\ell(q, t)$ for the context-assisted incremental Bayes value of a candidate with error rate q and coupling coordinate t in the binary family constructed below; t_m^0 is its coordinate at the common tie.

Theorem 5 (Fixed-score quantitative refinement) *For every finite $M \ge 2$ and every fixed regular strictly proper loss for a binary outcome, there exist a release and a full-support tie completion, possibly depending on that loss, such that every strict context-assisted ranking occurs in every sufficiently small TV neighborhood of radius η with a common adjacent value gap $g_\ell(\eta) > 0$. This gap may shrink as $\eta \downarrow 0$. If the coordinate paths are continuously differentiable at the tie and*

$$d_{\min} = \min_m \left|\partial_t v_\ell(q_m, t_m^0)\right| > 0, \qquad A_M = \left\lfloor M^2/4 \right\rfloor,$$

then, after reducing η_0, the common gap can be at least $d_{\min}\eta/(2A_M)$. For log loss the construction can ensure $d_{\min} \ge \kappa/2$ for the positive derivative coefficient κ computed below, giving gap at least $\kappa\eta/(4A_M)$. Consequently the coarser outcome-paired release $\mathcal{O}_{\mathrm{pair}}$ also shatters.

Proof. Let the binary outcome be a fair bit Y. Fix $0 < r < q_0 < 1/2$ and choose a compact neighbourhood $\mathcal{U} \subset (r, 1/2)$ of q_0. The distinct rates $q_m \in \mathcal{U}$ will be selected only after the common value range is fixed. Let $E_1, \dots, E_M$ be mutually independent Bernoulli errors, independent of Y, with rates q_m, and put $Z_m = Y \oplus E_m$. This fixes the complete law $P(Y, Z_1, \dots, Z_M)$, not only its candidate-wise margins.

For a vector $\theta \in [0, \bar\theta]^M$, define a context-error bit F conditionally on the entire candidate-error vector by

$$P_\theta(F = 1 \mid E_{1:M} = e) = r + \sum_{m=1}^{M} \theta_m(e_m - q_m), \tag{12}$$

and put $C = Y \oplus F$. Choose $\bar\theta > 0$ small enough that the right-hand side lies strictly between zero and one for every e and every $\theta \in [0, \bar\theta]^M$. This can be done uniformly over $q_m \in \mathcal{U}$, since $|\sum_m \theta_m(e_m - q_m)| \le M\bar\theta$. Centering and independence give

$$P_\theta(F = 1) = r, \qquad \mathrm{Cov}_\theta(F, E_m) = \theta_m q_m(1 - q_m).$$

Therefore $P(Y, C)$ is the same BSC(r) law throughout the box, while $P(Y, Z_{1:M})$ is unchanged because it was generated before F. Moreover, the exact pair law

$$P_\theta(F = 1, E_m = e) = P(E_m = e)\{r + \theta_m(e - q_m)\}, \qquad e \in \{0, 1\}, \tag{13}$$

depends on θ only through coordinate θ_m. Consequently the triple law of (Y, C, Z_m) also depends only on that coordinate. Hence the incremental value has the coordinatewise form $V_m^*(\theta) = v_\ell(q_m, \theta_m)$.

Release invariance and locality. Let P_0 denote the completion at $\theta = 0$. The maps from $(Y, E_{1:M}, F)$ to $(Y, C, Z_{1:M})$ are bijective for fixed Y, so direct summation of the two Bernoulli-mass changes gives, under the convention $\mathrm{TV}(P, Q) = \frac{1}{2}\sum_x |P(x) - Q(x)|$,

$$\mathrm{TV}(P_\theta, P_0) = \mathbb{E}\left|\sum_{m=1}^{M} \theta_m(E_m - q_m)\right| \le \sum_{m=1}^{M} 2\theta_m q_m(1 - q_m) \le \frac{M\bar{\theta}}{2}.$$

Thus the entire Cartesian family below is local to one reference completion; no distant-coupling construction is required.

A strict coordinatewise experiment order. It remains to show that each coordinate has a non-degenerate value range. Fix candidate m, write $q = q_m$, and let P_u denote the completion with $\theta_m = u$ while the other coordinates are held fixed. The additional binary observation $W_u = C \oplus Z_m = F \oplus E_m$ has channel

$$P_u(W_u = 1 \mid F = 0) = q - \frac{uq(1-q)}{1-r},$$
$$P_u(W_u = 1 \mid F = 1) = 1 - q - \frac{uq(1-q)}{r}.$$

The two rows coincide at

$$\theta^*(q) = \frac{r(1-r)(1-2q)}{q(1-q)(1-2r)} > 0.$$

For $0 \le s < t \le \theta^*(q)$, set $\lambda = (\theta^*(q) - t)/(\theta^*(q) - s) \in [0, 1)$. Entrywise, the channel satisfies $Q_t = \lambda Q_s + (1 - \lambda)Q_{\theta^*(q)}$, where $Q_{\theta^*(q)}$ has identical rows. Hence one can generate W_t by retaining the output of the s-channel with probability λ and otherwise replacing it by an independent Bernoulli draw having the common row distribution of $Q_{\theta^*(q)}$. Explicitly, if that common success probability is p_*, the post-processing kernel is

$$K_{s,t}(w' \mid w) = \lambda \mathbf{1}\{w' = w\} + (1 - \lambda)p_*^{w'}(1 - p_*)^{1-w'}, \qquad Q_t = Q_s K_{s,t}.$$

The common success probability at $\theta^*(q)$ is $(q - r)/(1 - 2r) \in (0, 1)$. To compare the experiments for predicting Y directly, define $X_u = (C, W_u)$. Retain C and apply $K_{s,t}$ only to W_s. Since $Q_t = Q_s K_{s,t}$ and $P(Y, C)$ is invariant, the resulting law is exactly $P_t(Y, X_t)$ and gives the Markov chain

$$Y \longrightarrow X_s \longrightarrow X_t.$$

Moreover, $Z_m = C \oplus W_u$, so (C, W_u) and (C, Z_m) generate the same σ-field under P_u. For $s < t \le \theta^*(q)$, the rows of Q_s are distinct and the replacement kernel assigns positive probability to both outputs. Conditional on either $C = c$, the two posteriors $P_s(Y \mid C = c, W_s = 0)$ and $P_s(Y \mid C = c, W_s = 1)$ are therefore distinct, while each $P_t(Y \mid C = c, W_t = w')$ is the mixture

$$\sum_{w \in \{0,1\}} P_s(Y \mid C = c, W_s = w)\, P_s(W_s = w \mid C = c, W_t = w'),$$

with both weights strictly positive. Strict Jensen for the strictly concave Bayes risk H_ℓ gives $R^*_{\ell,P_t}(X_t) > R^*_{\ell,P_s}(X_s)$. Since $R^*_\ell(C)$ is invariant and $R^*_{\ell,P_u}(X_u) = R^*_{\ell,P_u}(C, Z_m)$, it follows that

$$v_\ell(q, s) > v_\ell(q, t) \quad \text{for} \quad 0 \le s < t \le \theta^*(q).$$

This direct comparison predicts Y throughout and does not assume label symmetry. There is no boundary exception: the initial strict probability-box choice, $q \in (r, 1/2)$, and $p_* \in (0, 1)$ give positive mass to every posterior mixed above. Because $\theta^*(q)$ is continuous and positive on compact $\mathcal{U}$, shrink $\bar{\theta}$ further so that $\bar{\theta} < \inf_{q \in \mathcal{U}} \theta^*(q)$.

One common range with a strict marginal order. If all $q_m = q_0$, the coordinate ranges $[v_\ell(q_0, \bar{\theta}), v_\ell(q_0, 0)]$ coincide and have positive width. Choose a closed positive-width interval strictly inside this common range. On every compact (q, t) rectangle inside the strict probability box, $v_\ell(q, t)$ is jointly continuous: all relevant cell probabilities are continuous, and the Bayes-risk perspective remains continuous at zero-mass cells. Hence, after shrinking $\mathcal{U}$ around q_0, this interval remains inside every coordinate range for every $q \in \mathcal{U}$. We may therefore choose distinct $q_1 < \cdots < q_M$ in the same $\mathcal{U}$. Their marginal BSC values satisfy $U_1^* > \cdots > U_M^* > 0$: moving the error rate toward one half is a nontrivial Blackwell garbling, and strict concavity again makes Bayes risk strictly increase Blackwell (1953); Gneiting & Raftery (2007).

One fixed-release tie completion at every local radius. Keep the already selected distinct $q_1 < \cdots < q_M$, the fixed $\bar{\theta}$, and the resulting release $\widetilde{\mathcal{O}}$ fixed. Choose v_0 in the interior of the common coordinate interval. Strict monotonicity and continuity give unique $t_m^0 \in (0, \bar{\theta})$ satisfying

$$v_\ell(q_m, t_m^0) = v_0, \qquad m = 1, \ldots, M.$$

Let $Q_0 = P_{t^0}$ be this fixed-release tie completion. For each coordinate, the inverse of $v_\ell(q_m, \cdot)$ is continuous at v_0. Consequently, for every sufficiently small $\eta > 0$, one may choose M distinct values arbitrarily close to v_0, assign them in any requested order, and take their coordinatewise inverse images so that the resulting P_θ lies in $B_{\mathrm{TV}}(Q_0, \eta)$. The same $\widetilde{\mathcal{O}}$, Q_0, and strict marginal order work for every such radius. Because there are finitely many coordinates and assigned values, every radius has a common positive adjacent gap.

Suppose additionally that the coordinate paths are continuously differentiable at the tie and let

$$d_{\min} = \min_m |\partial_t v_\ell(q_m, t_m^0)| > 0.$$

After shrinking a common neighborhood, the inverse maps satisfy $|t_m(v) - t_m(v_0)| \le (2/d_{\min})|v - v_0|$. Let

$$a_j = (M + 1 - 2j)/2, \qquad A_M = \sum_{j=1}^{M} |a_j| = \lfloor M^2/4 \rfloor.$$

For a requested best-to-worst permutation π, target the centered values $v_0 + g a_j$ for candidate $\pi(j)$. Their adjacent gap is g, and the inverse bound plus the centered-coupling TV identity gives

$$\mathrm{TV}(P_\theta, Q_0) \le \frac{1}{2} \sum_m |\theta_m - t_m^0| \le \frac{g A_M}{d_{\min}}.$$

Thus $g = d_{\min}\eta/(2A_M)$ lies strictly inside the TV ball for all sufficiently small η and proves the quantitative fixed-release clause.

For completeness, a radius-dependent auxiliary construction supplies a score-general explicit modulus without differentiability. Set $\bar{\theta} = \eta/M$ and write $\omega_\ell(\eta/M) = v_\ell(q_0, 0) - v_\ell(q_0, \eta/M)$. Joint continuity permits distinct nearby $q_m(\eta)$ whose ranges share an interval of width $\omega_\ell(\eta/M)/2$, giving adjacent gap $\omega_\ell(\eta/M)/[2(M-1)]$. This auxiliary release may vary with η; it is not needed for the stronger fixed-release topological statement.

Log-loss coefficient and coherent gluing. For log loss, $v_{\log}(q, \theta) = I(F; W)$ in this binary construction. Indeed, because Y is fair and independent of (F, E), $C = Y \oplus F$ is independent of (F, W), so $I(Y; Z_m \mid C) = I(F; W \mid C) = I(F; W)$. Put $p_0 = P_{\theta=0}(W = 1) = q_0 + r - 2q_0 r < 1/2$. Differentiating the binary-entropy expression

$$I(F; W) = h(p_\theta) - (1 - r)h(a_\theta) - r h(b_\theta),$$

where $a_\theta = q_0 - \theta q_0(1 - q_0)/(1 - r)$, $b_\theta = 1 - q_0 - \theta q_0(1 - q_0)/r$, and $p_\theta = p_0 - 2\theta q_0(1 - q_0)$. Since $h'(1 - q_0) = -h'(q_0)$, the two conditional-entropy derivatives cancel at zero, giving

$$-\left.\frac{\partial v_{\log}(q_0, \theta)}{\partial \theta}\right|_{\theta = 0^+} = 2q_0(1 - q_0) \log \frac{1 - p_0}{p_0} =: \kappa > 0.$$

Hence $\omega_{\log}(t) = \kappa t + o(t)$. By taking $\bar{\theta}$ and the q-neighborhood sufficiently small before fixing the release, continuity of the log-loss derivative gives $d_{\min} \ge \kappa/2$ at the tie completion. The fixed-release gap above is therefore at least $\kappa\eta/(4A_M)$. The auxiliary radius-adapted modulus remains $\kappa\eta/[2M(M-1)] + o(\eta)$.

The common-interval inverse construction above realizes each requested order with a positive adjacent gap. Marginalizing $\widetilde{\mathcal{O}}$ shows that the coarser release $\mathcal{O}_{\mathrm{pair}}$ also shatters. □

This quantitative refinement permits the selected hidden couplings and nearby error rates to depend on the fixed loss. It supplies no numerical gap uniform over scores; the score-uniform ordinal statement is instead Theorem 1.

Theorem 8 (Centered tangent criterion for structural local shattering) *Let a full-support finite-alphabet completion Q_0 lie in a fiber of a linear release map $\mathcal{R}$, let $F = (F_1, \dots, F_M)$ be continuously differentiable near Q_0 within that fiber, and suppose $F(Q_0) = v_0\mathbf{1}$. Write $T = \ker D\mathcal{R}(Q_0)$ and $\Pi = I - M^{-1}\mathbf{1}\mathbf{1}^\top$. If $\operatorname{rank}(\Pi DF(Q_0)|_T) = M - 1$, then the fixed release $\mathcal{R}(Q_0)$ realizes every strict ranking in every sufficiently small TV neighborhood of Q_0. Within a connected open finite-dimensional real-analytic family, if a maximal minor of this centered derivative is nonzero somewhere, the rank-deficient parameter set has Lebesgue measure zero.*

D.2 Proof of the centered tangent criterion

Proof of Theorem 8. Let N_{atom} be the number of coordinates in the joint probability vector (not the empirical row count). Use affine Euclidean coordinates for the finite probability simplex. Full support makes a neighborhood of Q_0 open in its affine hull, and linearity of $\mathcal{R}$ makes its local fiber exactly $Q_0 + T$, where $T = \ker D\mathcal{R}(Q_0)$. Restrict the centered value map to that fiber:

$$G(h) = \Pi\{F(Q_0 + h) - F(Q_0)\}, \qquad h \in T.$$

The assumed rank says that $A = DG(0) : T \to \mathbf{1}^\perp$ is surjective. The submersion theorem therefore supplies a continuously differentiable local right section s on a neighborhood of zero in $\mathbf{1}^\perp$ such that $G\{s(w)\} = w$ and $s(0) = 0$. Choose the section with $Ds(0) = A^+$, the Moore–Penrose right inverse. If σ_+ is the smallest nonzero singular value of A, continuity permits the neighborhood to be shrunk until

$$\|s(w)\|_2 \le \frac{2}{\sigma_+}\|w\|_2.$$

For a best-to-worst permutation π, assign its candidates the centered rank vector with entries $a_j = (M + 1 - 2j)/2$. It lies in $\mathbf{1}^\perp$, has adjacent gap one, and has norm

$$B_M = \left(\sum_{j=1}^{M} a_j^2\right)^{1/2} = \sqrt{M(M^2-1)/12}.$$

Take $w_\pi = g a_\pi$ with $g = \sigma_+\eta/(2\sqrt{N_{\mathrm{atom}}}B_M)$. For all sufficiently small η, every w_π lies in the common domain of the right section and $Q_0 + s(w_\pi)$ remains a probability law in the same release fiber. Moreover,

$$\mathrm{TV}\{Q_0 + s(w_\pi), Q_0\} = \frac{1}{2}\|s(w_\pi)\|_1 \le \frac{\sqrt{N_{\mathrm{atom}}}}{2}\|s(w_\pi)\|_2 \le \frac{\sqrt{N_{\mathrm{atom}}}}{\sigma_+}\|w_\pi\|_2 = \frac{\eta}{2} < \eta.$$

Its centered value vector is w_π; adding an unspecified common component does not change order, so its adjacent gap is g. There are finitely many permutations, hence one sufficiently small radius works for all of them.

Surjectivity is open because a nonzero maximal minor remains nonzero under a small perturbation. For the stated analytic qualification, choose fixed orthonormal bases of T and $\mathbf{1}^\perp$. If one maximal minor of the centered derivative matrix is nonzero somewhere on a connected open finite-dimensional real-analytic family, that minor is not the zero analytic function. The rank-deficient locus is contained in its zero set, which has Lebesgue measure zero. Without these analytic and nonvanishing-minor premises, no measure-zero genericity conclusion is asserted.

Finally, in the differentiable instance of Theorem 5, the θ directions lie in T and the coordinate-separation identity makes the uncentered Jacobian diagonal with entries $\partial_t v_\ell(q_m, t_m^0)$. If all are nonzero, multiplication by Π leaves rank $M - 1$, so the construction is an explicit witness to the criterion. □

D.3 Extension: overlapping smooth-density calibrated-forecast lift

Lemma 4 (Bayes-risk continuity under total variation) *Let Y take values in a finite set and let X take values in an arbitrary standard Borel space. If H is continuous on the outcome simplex, then*

$$\mathfrak{R}_H(P) := \mathbb{E}_P H\{P(Y = \cdot \mid X)\}$$

is continuous as a functional of the joint law $P_{Y,X}$ in total variation. The assertion also holds with X replaced by (C, X) for finite C.

Proof. For two joint laws P, Q, let $\mu = P_X + Q_X$ and write $p_y = dP(Y = y, X \in \cdot)/d\mu$ and $q_y = dQ(Y = y, X \in \cdot)/d\mu$. Define the perspective

$$\Phi(a) = \begin{cases} (\sum_y a_y) H\Big(a/\sum_y a_y\Big), & \sum_y a_y > 0, \\ 0, & a = 0. \end{cases}$$

Continuity of H on the compact simplex makes Φ continuous, uniformly continuous, and bounded on the compact cube $[0, 1]^{|\mathcal{Y}|}$. Moreover,

$$\mathfrak{R}_H(P) = \int \Phi\{p(x)\}\, d\mu(x), \qquad \mathfrak{R}_H(Q) = \int \Phi\{q(x)\}\, d\mu(x),$$

and

$$\int \|p(x) - q(x)\|_1\, d\mu(x) = 2\,\mathrm{TV}(P, Q).$$

Let $B = \sup|\Phi|$ and let ω_Φ be a modulus of continuity on this cube. For any $\delta > 0$, split the domain into $A_\delta = \{\|p - q\|_1 \le \delta\}$ and its complement. Since $\mu(X) = 2$ and $\mu(A_\delta^c) \le \delta^{-1} \int \|p - q\|_1 d\mu$,

$$|\mathfrak{R}_H(P) - \mathfrak{R}_H(Q)| \le 2\omega_\Phi(\delta) + \frac{2B}{\delta} \int \|p - q\|_1 d\mu.$$

First choose δ so that the first term is small and then let the total variation tend to zero. Treating (C, X) as one signal proves the last claim. □

Corollary 5 (Fixed overlapping calibrated-forecast lift) *For every $K \ge 2$, finite M, and fixed regular strictly proper loss for a K-class outcome, the construction in Theorem 5 has a full-support, exactly calibrated-forecast lift that preserves its fixed-release local all-order conclusion. The lift and tie may depend on the fixed loss. Each forecast is interior, has a smooth density positive throughout the simplex interior, and has unique argmax almost surely; the latent finite atom is not exactly decodable. This corollary does not claim one Blackwell-ordered lift shared across losses. Separately, any declared finite family of calibrated-atom witness worlds with a common positive adjacent gap G admits one common overlapping smoothing kernel that preserves all its strict orders with gap at least $G/2$.*

Proof of Corollary 5. Fix the stated K-class loss. Write the binary theorem's outcome as $Y_0 \in \{1, 2\}$ and its candidate code as $J_m := Z_m$. For $K = 2$, set $S_m(j) = P(Y_0 = \cdot \mid J_m = j)$. The BSC rates lie strictly below $1/2$, so the two posterior vectors are distinct, interior, and have unique argmax j. They are calibrated by construction, and because the map $j \mapsto S_m(j)$ is injective, conditioning on S_m and on J_m gives the same Bayes risk. Direct evaluation of the forecast under the proper score also attains this marginal Bayes risk, $\mathbb{E}\ell(Y, S_m) = R^*_\ell(S_m) = R^*_\ell(J_m)$; context-assisted Bayes evaluation instead uses $R^*_\ell(C, S_m)$.

For $K > 2$, apply to Y_0 a fixed channel A_ε with

$$A_\varepsilon(k \mid b) = \begin{cases} 1 - (K - 1)\varepsilon, & k = b, \\ \varepsilon, & k \ne b, \end{cases} \qquad 0 < \varepsilon < 1/K,$$

and call the resulting outcome Y. The channel is independent of $(C, J_{1:M})$ given Y_0 and is the same in every hidden world. Therefore the already fixed $P(Y_0, C)$ and $P(Y_0, J_{1:M})$ determine fixed releases $P(Y, C)$ and $P(Y, J_{1:M})$. Define the released candidate forecast by

$$S_m(j) := P(Y = \cdot \mid J_m = j).$$

Every coordinate is positive. For sufficiently small ε, the two vectors are distinct and retain unique argmax j, hence $j \mapsto S_m(j)$ is injective. Consequently $P(Y = \cdot \mid S_m) = S_m$ exactly, direct proper-score risk equals the marginal Bayes risk conditional on J_m, and $P(Y, S_{1:M})$ is a deterministic pushforward of the fixed $P(Y, J_{1:M})$.

Work over the compact coordinate box used in the fixed-release proof. At $\varepsilon = 0$, its marginal and context-assisted Bayes risks reduce to those of the binary construction (with zero mass on the added classes). Continuity of the regular proper score's Bayes risk, compactness of the box, and finiteness of the candidates give uniform convergence as $\varepsilon \downarrow 0$. Hence one fixed positive ε preserves the strict marginal order and leaves a closed positive-width interval $[v_-, v_+]$ strictly inside every atom-lifted coordinate range, with $v_m^\varepsilon(0) > v_+ > v_- > v_m^\varepsilon(\bar{\theta})$ for every m. The channel is fixed before any local radius is quantified and cannot increase total variation. Denote the resulting finite forecast atoms by A_m and now smooth them. For $K = 2$, the same argument starts directly from the two interior posterior atoms and omits A_ε.

Let $u = (1/K, \ldots, 1/K)$ be the simplex barycenter. On the simplex affine hull, normalize the permutation-symmetric reference density

$$h(s) \;\propto\; \begin{cases} \exp\Big\{-1/\prod_{j=1}^K s_j\Big\}, & s \in \operatorname{relint} \Delta_K, \\ 0, & \text{otherwise.} \end{cases}$$

It is positive throughout $\operatorname{relint} \Delta_K$, its zero extension is C^∞: in any affine coordinate chart, every interior derivative is a finite sum of bounded smooth coefficients times $p(s)^{-N} \exp\{-1/p(s)\}$, where $p(s) = \prod_j s_j$; all such terms tend to zero as $p(s) \downarrow 0$. Thus every derivative extends by zero at the boundary. Permutation symmetry then gives $\int s h(s) d\lambda_{K-1}(s) = u$. Let $a_{\min} = \min_{m,a,y} a_y > 0$. Choose a common tail weight $0 < \tau < \min\{1, K a_{\min}\}$ and recenter every atom at

$$\widetilde{a} = \frac{a - \tau u}{1 - \tau} \in \operatorname{relint} \Delta_K.$$

Coordinate differences are merely rescaled, so $\widetilde{a}$ retains the unique argmax of a. Finiteness gives one $r_*(\tau) > 0$ such that, for every candidate, all tangent balls $B_{m,a}(r)$ of radius $0 < r < r_*(\tau)$ around its recentered atoms are pairwise disjoint, lie in the simplex interior, and stay inside their corresponding argmax cells. Because $\widetilde{a} \to a$ as $\tau \downarrow 0$, $r_*(\tau)$ is bounded away from zero for all sufficiently small τ. Let $b_{m,a,r}$ be a normalized radially symmetric C^∞ bump supported inside $B_{m,a}(r)$, positive on a smaller concentric ball, and vanishing to all orders at the support boundary. Tangent-space symmetry gives $\int s b_{m,a,r}(s) d\lambda_{K-1}(s) = \widetilde{a}$. Define

$$g_{m,a}^{\tau,r}(s) = (1 - \tau) b_{m,a,r}(s) + \tau h(s). \tag{14}$$

Then $g_{m,a}^{\tau,r}$ is a globally C^∞ density, is positive throughout the simplex interior, and has mean exactly $(1 - \tau)\widetilde{a} + \tau u = a$. Thus all atom-conditional densities have the same full interior support even though their high-mass cores are disjoint. For $a_y > 0$, define the probability kernel

$$K_m(ds \mid Y = y, A_m = a) = \frac{s_y}{a_y} g_{m,a}^{\tau,r}(s)\, d\lambda_{K-1}(s). \tag{15}$$

It is nonnegative and, for every (y, a), $\int (s_y/a_y) g_{m,a}^{\tau,r}(s) d\lambda_{K-1}(s) = a_y/a_y = 1$. Apply these kernels independently across candidates conditional on $(Y, A_{1:M})$, using the same kernels in every hidden world, and call the new outputs $S_{1:M}$.

This conditional kernel is only a disintegration used to specify the joint law; it is not a deployment algorithm that observes the realized outcome. Equivalently, disintegrate the resulting law in the signal-first direction: draw $(C, S_{1:M})$ from its marginal and then draw Y from $P(Y \mid C, S_{1:M})$. Both orders generate the same observable joint law and all claims below concern that law rather than a causal direction.

Exact calibration does not require disjoint support. Since $P(Y = y, A_m = a) = P(A_m = a) a_y$, summing over all atoms gives

$$dP(Y = y, S_m = s) = s_y \sum_a P(A_m = a) g_{m,a}^{\tau,r}(s)\, d\lambda_{K-1}(s).$$

Hence the forecast density is $\sum_a P(A_m = a) g_{m,a}^{\tau,r}(s)$ and $P(Y = y \mid S_m = s) = s_y$ almost everywhere. Calibration is exact. The marginal density is globally smooth and positive throughout

the simplex interior. For any two positive-mass atoms, their conditional laws given any outcome have positive densities on that same interior and are mutually absolutely continuous. Therefore no measurable decoder can recover A_m from S_m with probability one. Tie hyperplanes have affine Lebesgue measure zero, so the forecast has unique argmax almost surely, and direct proper-score risk equals the marginal Bayes risk $R^*_\ell(S_m)$. Context-assisted Bayes evaluation uses $R^*_\ell(C, S_m)$. Moreover, $P(Y, S_{1:M})$ is the image of the already fixed $P(Y, A_{1:M})$ under one fixed product kernel, so the complete joint forecast release remains identical in all ranking worlds. The data-processing inequality for total variation preserves the local radius bound.

It remains to preserve the strict values without exact atom decoding. Compare the calibrated kernel above with the outcome-independent channel $\overline{K}_m(ds \mid A_m = a) = g^{\tau,r}_{m,a}(s)d\lambda_{K-1}(s)$. Their conditional total variation is at most

$$\mathrm{TV}\big(K_m(\cdot \mid y, a), \overline{K}_m(\cdot \mid a)\big) = \frac{1}{2}\int \left|\frac{s_y}{a_y} - 1\right| g^{\tau,r}_{m,a}(s)d\lambda_{K-1}(s) \tag{16}$$

$$\leq \frac{1}{2a_{\min}}\left(r + \frac{\tau}{1-\tau} + \tau\right) = O(r+\tau). \tag{17}$$

Here the core contribution uses $|s_y - \widetilde{a}_y| \leq r$ and $|\widetilde{a}_y - a_y| \leq \tau/(1-\tau)$, while the tail contribution is at most τ. The bound is uniform over the finite candidates, atoms, outcomes, and the entire compact coordinate box. Averaging it shows that each calibrated joint law of (Y, C, S_m) is $O(r+\tau)$ in TV from its comparison counterpart.

Under the comparison channel, decode a when S_m falls in its disjoint core ball. The only errors come from the common-tail mixture, so a decoder $D_m(S_m)$ satisfies

$$P\{D_m(S_m) \neq A_m\} \leq \tau. \tag{18}$$

Under this comparison law,

$$Y \longrightarrow A_m \longrightarrow S_m \longrightarrow D_m(S_m)$$

and

$$Y \longrightarrow (C, A_m) \longrightarrow (C, S_m) \longrightarrow (C, D_m(S_m)).$$

Blackwell monotonicity therefore squeezes the comparison risks:

$$R^*(A_m) \leq R^*(S_m) \leq R^*\{D_m(S_m)\},$$
$$R^*(C, A_m) \leq R^*(C, S_m) \leq R^*\{C, D_m(S_m)\}.$$

The coupling in Equation (18) puts the joint laws of $(Y, D_m(S_m))$ and (Y, A_m), and likewise those with C adjoined, within TV at most τ. Lemma 4 makes the outer risks converge to their atom counterparts. The squeeze yields, under the comparison law,

$$R^*(S_m) \to R^*(A_m), \qquad R^*(C, S_m) \to R^*(C, A_m)$$

as $(\tau, r) \to (0, 0)$ with $r < r_*(\tau)$. Equation (17) and the same lemma transfer the convergence to the calibrated law.

The convergence just proved is uniform over the full compact coordinate box: the comparison bound is independent of the hidden coordinate, the decoder bound is always at most τ, and Lemma 4 supplies one common modulus. Choose one positive τ and then one $0 < r < r_*(\tau)$ so that the smoothed marginal order remains strict and, writing the smoothed value path of candidate m as $w_m(t)$,

$$w_m(0) > v_+ > v_- > w_m(\bar{\theta}), \qquad m = 1, \ldots, M.$$

These parameters and the resulting product kernel are now fixed before the TV radius is quantified.

For fixed m, the joint density of (Y, C, S_m) is affine in t. The perspective of the concave Bayes risk is concave, so $R^*(C, S_m)$ is concave in t and $w_m(t) = R^*(C) - R^*(C, S_m)$ is convex. Fix any $v_0 \in (v_-, v_+)$. The sublevel set $\{t : w_m(t) \leq v_0\}$ is an interval containing $\bar{\theta}$ but not zero. Its left endpoint $t^0_m \in (0, \bar{\theta})$ satisfies $w_m(t^0_m) = v_0$. Because $w_m(\bar{\theta}) < v_0$, convexity excludes a right-hand plateau at v_0. Choose $u_m > t^0_m$ arbitrarily close with $w_m(u_m) < v_0$ and put

$$c_m = \frac{v_0 - w_m(u_m)}{u_m - t^0_m} > 0, \qquad c_{\mathrm{sm}} = \min_m c_m.$$

Monotonicity of convex secant slopes gives, after restricting to a common neighborhood of the crossings,

$$|w_m(t) - v_0| \geq c_m |t - t_m^0|.$$

Thus each path has a continuous local inverse there with $|t - t_m^0| \leq |w_m(t) - v_0|/c_{\mathrm{sm}}$.

Apply the already fixed product kernel to the completion at $t^0 = (t_1^0, \ldots, t_M^0)$ and call the result Q_0^{sm}. Its complete $P(Y, S_{1:M})$ and $P(Y, C)$ define one fixed smooth release, and all its candidate values equal v_0. For a requested best-to-worst permutation, use the centered rank levels $a_j = (M+1-2j)/2$ and target $v_0 + g a_j$, where

$$g = \frac{c_{\mathrm{sm}} \eta}{2 A_M}.$$

For all sufficiently small η, the local inverse branches attain these values. Data processing and the finite coordinate-box TV identity give

$$\mathrm{TV}(Q_t^{\mathrm{sm}}, Q_0^{\mathrm{sm}}) \leq \frac{1}{2} \sum_m |t_m - t_m^0| \leq \frac{g A_M}{2 c_{\mathrm{sm}}} = \frac{\eta}{4} < \eta.$$

The adjacent gap is g, proving the fixed-channel, all-radii clause.

For the separate finite-family fidelity clause, start from any declared finite calibrated-atom witness family with common gap $G > 0$. Choose one common $\tau > 0$ and then one common $0 < r < r_*(\tau)$ so that every relevant marginal or context-assisted value changes by less than $G/4$. Each adjacent difference then decreases by less than $G/2$, so all strict orders persist and the final gap is at least $G/2$. If this family was obtained from the binary witness at a declared radius, first choosing the outcome channel to retain half the binary gap leaves at least one quarter of the binary guarantee. In particular the log-loss gap is at least $\kappa\eta/[8M(M-1)] + o(\eta)$ for that fidelity choice. This proves all claims. The common positive tail removes exact finite-atom decodability, but the construction remains an existence theorem, not native empirical calibration, natural-geometry typicality, or prevalence. □

D.4 Extension: auxiliary candidate-wise release results

Here the weaker release is $\mathcal{O}_{\mathrm{pair}} = \{P(Y, C), P(Y, Z_1), \ldots, P(Y, Z_M)\}$; in this subsection and its proofs, $\mathcal{O}$ abbreviates $\mathcal{O}_{\mathrm{pair}}$. For candidate m, let $\Gamma_m(\mathcal{O})$ be the compatible triple-law fiber and $\mathcal{I}_m(\mathcal{O}) = \{V_m^*(Q_m) : Q_m \in \Gamma_m(\mathcal{O})\}$. Let $\mathcal{V}(\mathcal{O})$ be the value vectors attainable by full laws preserving $\mathcal{O}$. For a best-to-worst permutation π, define

$$\gamma_\pi(\mathcal{O}) = \max_{v \in \mathcal{V}(\mathcal{O})} \min_{j<M} \{v_{\pi(j)} - v_{\pi(j+1)}\},$$

$$\gamma_{\mathrm{worst}}(\mathcal{O}) = \min_{\pi \in S_M} \gamma_\pi(\mathcal{O}), \qquad \rho(\mathcal{O}) = \max\{0, \gamma_{\mathrm{worst}}(\mathcal{O})\}.$$

The signed margin records an impossible order when negative; ρ records the nonnegative worst-order gap. For finite alphabets, each $\Gamma_m(\mathcal{O})$ is a nonempty compact convex polytope and V_m^* is continuous on it; hence $\mathcal{I}_m = [L_m, U_m]$ is a compact interval. Let $\Gamma(\mathcal{O})$ contain all full laws preserving the release.

Theorem 9 (Exact Cartesian identification) *For finite alphabets, $\mathcal{V}(\mathcal{O}) = \mathcal{I}_1 \times \cdots \times \mathcal{I}_M$. Thus every candidate-wise selection of attainable values is jointly realizable in one coherent M-candidate world.*

Let a raw-marginal schema $\mathcal{H}$ release exact tables $\{P(X_S) : S \in \mathcal{H}\}$ for scopes $S \subseteq \{Y, C, Z_1, \ldots, Z_M\}$.

Theorem 10 (Raw-marginal disclosure frontier) *For finite Y, C, Z_m, each with at least two symbols, and any regular strictly proper loss on Y, $V_m^*(C)$ is universally point-identified from $\mathcal{H}$ if and only if some $S \in \mathcal{H}$ contains $\{Y, C, Z_m\}$. The whole value vector is universally identified if and only if this holds for every m.*

Theorem 11 (Exact interval rank shattering) *Let $\Delta = \min_{a \neq b}(U_a - L_b)$. Then*

$$\gamma_{\mathrm{worst}}(\mathcal{O}) = \begin{cases} \Delta/(M-1), & \Delta \geq 0, \\ \Delta, & \Delta < 0. \end{cases}$$

For $\epsilon \geq 0$, the same release is compatible with all $M!$ rankings, each with adjacent gaps larger than ϵ, if and only if $\Delta > (M-1)\epsilon$.

D.5 ATTAINABLE VALUES ARE INTERVALS

Proof. For finite alphabets, $\Gamma_m(\mathcal{O})$ is the intersection of a probability simplex with finitely many affine marginal constraints. It is therefore a nonempty compact convex polytope. To make continuity at zero-mass cells explicit, for a nonnegative vector $a = (a_y)_{y\in\mathcal{Y}}$ define the perspective

$$\phi_\ell(a) = \begin{cases} (\sum_y a_y) H_\ell\Big(a/(\sum_y a_y)\Big), & \sum_y a_y > 0, \\ 0, & a = 0. \end{cases}$$

Because H_ℓ is continuous and therefore bounded on the compact simplex, $\phi_\ell(a) \to 0$ whenever $\sum_y a_y \to 0$. Thus ϕ_ℓ is continuous on the nonnegative orthant, and every Bayes risk here is a finite sum of such terms. Hence $Q_m \mapsto V_m^*(Q_m)$ is continuous, including on faces with zero mass. A convex set is connected, and the continuous image of a connected compact set in $\mathbb{R}$ is a compact interval. Therefore $\mathcal{I}_m = [L_m, U_m]$. □

No binary-support assumption is used here. Binary support is needed only by our current exact two-coordinate optimizer.

D.6 PROOF OF THEOREM 9

Proof. Every full law in $\Gamma(\mathcal{O})$ projects to a compatible candidate triple, so its value vector lies in $\prod_m \mathcal{I}_m$.

Conversely, fix any vector $(v_1, \dots, v_M) \in \prod_m \mathcal{I}_m$. The interval lemma supplies, for every candidate m, a compatible triple law $Q_m(Y, C, Z_m)$ attaining v_m. All selected triples share the released $P(Y, C)$. Define

$$Q(y, c, z_1, \dots, z_M) = P(y, c) \prod_{m=1}^{M} Q_m(z_m \mid y, c).$$

When $P(y, c) = 0$, choose the conditional factors arbitrarily; the corresponding full cell remains zero. Marginalizing Q over every candidate except m recovers Q_m, so $Q \in \Gamma(\mathcal{O})$ and its value vector is exactly $(v_1, \dots, v_M)$. This proves the reverse inclusion. □

D.7 PROOF OF THE RAW-MARGINAL DISCLOSURE FRONTIER

Proof of Theorem 10. If some released scope S contains $\{Y, C, Z_m\}$, marginalizing $P(X_S)$ identifies $P(Y, C, Z_m)$ and therefore both terms in $V_m^*(C) = R_\ell^*(C) - R_\ell^*(C, Z_m)$. This proves sufficiency.

For necessity, suppose no released scope contains all three variables. Choose two symbols from each alphabet, identify them with zero and one, and put zero mass on all remaining symbols. Make every remaining variable independent with the same law in both worlds. In world Q_0, let Y, C, Z_m be mutually independent fair bits on the selected symbols. In world $Q_\oplus$, draw C, Z_m independently and set $Y = C \oplus Z_m$. The univariate and all three pairwise marginals of (Y, C, Z_m) are identical in both worlds. Every released scope omits at least one of these three variables, so its law is one of those proper-subset marginals times the same independent law of the remaining variables. Hence the complete raw-marginal release is identical.

Let e_0, e_1 denote the two selected outcome vertices. In Q_0, observing Z_m after C has zero value. In $Q_\oplus$, Y has distribution $(e_0 + e_1)/2$ conditional on C alone but is deterministic given (C, Z_m), so

$$V_m^*(C) = H_\ell((e_0 + e_1)/2) - \tfrac{1}{2}\{H_\ell(e_0) + H_\ell(e_1)\} > 0,$$

where strict inequality follows from strict concavity of the Bayes risk along the segment joining e_0 and e_1. Thus V_m^* is not universally identified. Applying the condition to every candidate proves the vector statement. □

Proposition 4 (Fixed-rule interaction criterion and local shattering) *Fix a loss ℓ and a possibly randomized deployment rule $\mathcal{A}$ using only binary (C, Z). Define $h(y, c, z) = \mathbb{E}_U[\ell(y, \mathcal{A}(c, z; U))]$ and the within-outcome interaction contrast*

$$\delta_y = h(y, 0, 0) - h(y, 0, 1) - h(y, 1, 0) + h(y, 1, 1).$$

For candidate m, its population-risk identified set under $\mathcal{O}$ is exactly

$$\{\mathbb{E}_Q h(Y, C, Z_m) : Q \in \Gamma_m(\mathcal{O})\}.$$

The risk is identified from $P(Y, C)$ and $P(Y, Z_m)$ for every compatible release if and only if $\delta_y = 0$ for every y.

If $\delta_{y_0} \neq 0$ for some y_0, then for every finite $M \geq 2$ and every $\lambda > 0$ there is a release fixing $P(Y, C)$ and the complete $P(Y, Z_{1:M})$ with the following property. For all sufficiently small $\eta > 0$, all $M!$ strict rankings of the risks of the same fixed rule *$\mathcal{A}(C, Z_m)$ occur within TV distance η of one completion while*

$$1 \leq \mathrm{OR}(C, Z_m \mid Y = y) \leq e^{\lambda} \quad \textit{for every } m, y.$$

Writing $A_M = \sum_{j=1}^{M} |j - (M+1)/2| = \lfloor M^2/4 \rfloor$, the construction has adjacent risk gap $|\delta_{y_0}|\eta/(2A_M)$. Thus the gap is positive and first-order in η at fixed M, but it is not uniform in leaderboard size.

Proof. Risk is a linear functional of the compatible triple law. For fixed y, a 2×2 table functional is determined by its row and column margins exactly when its mixed difference is zero. Indeed, $\delta_y = 0$ gives

$$h(y, c, z) = h(y, c, 0) + h(y, 0, z) - h(y, 0, 0),$$

whose expectation is determined by $P(Y, C)$ and $P(Y, Z_m)$. Conversely, a nonzero mixed difference changes along the one-dimensional fixed-margin fiber.

For the simultaneous claim, fix a full-support law p for Y and let Q_0 make $C, Z_{1:M}$ independent fair bits given Y. For a vector $\theta = (\theta_1, \ldots, \theta_M)$ with $\|\theta\|_1 < 1$, define

$$Q_\theta(y, c, z_{1:M}) = Q_0(y, c, z_{1:M}) \left\{ 1 + \mathbf{1}\{y = y_0\} \sum_{m=1}^{M} \theta_m (-1)^{c + z_m} \right\}.$$

Every cell is positive. Summing over C removes all perturbations, as does summing over $Z_{1:M}$, so every Q_θ has the same $P(Y, C)$ and complete $P(Y, Z_{1:M})$. Terms indexed by $j \neq m$ also vanish from the (Y, C, Z_m) marginal, giving the exact risk identity

$$R_m(\theta) = R_0 + \frac{p(y_0)\delta_{y_0}}{4} \theta_m.$$

Choose a small common $\theta_0 > 0$. For a requested permutation π, assign

$$\theta_m = \theta_0 + \mathrm{sign}(\delta_{y_0}) s \{\pi(m) - (M+1)/2\}, \qquad s = \frac{2\eta}{p(y_0) A_M},$$

where $\pi(m) \in \{1, \ldots, M\}$ is candidate m's requested risk rank. For sufficiently small η, all coordinates remain positive and $\|\theta\|_1 < 1$. Relative to the common completion $Q_{(\theta_0, \ldots, \theta_0)}$,

$$\mathrm{TV} \leq \frac{p(y_0)}{2} s A_M = \eta,$$

and adjacent risks differ by $p(y_0)|\delta_{y_0}|s/4 = |\delta_{y_0}|\eta/(2A_M)$. Finally, at y_0 the candidate-wise conditional table is proportional to $1 + \theta_m(-1)^{c+z}$, so its odds ratio is $\{(1+\theta_m)/(1-\theta_m)\}^2$; at other outcomes it is one. Taking θ_0 and then η small enough puts every ratio in $[1, e^{\lambda}]$. This proves positive-association, bounded-odds-ratio local shattering. If a rule ignores C, or more generally has every $\delta_y = 0$, the first part shows why the construction cannot move its risk. □

A fixed binary logit pool is a concrete nondegenerate rule: report $\sigma\{(2C-1) + (2Z_m - 1)\}$ and score it by log loss. Its two interaction contrasts both equal 0.867562. The executable five-candidate audit realizes all 120 orders at four shrinking TV scales, with maximum release error 5.56×10^{-17} and all conditional log odds ratios below 0.105 in absolute value. This is a theorem falsification check, not prevalence evidence.

The XOR witness in the proof of Theorem 10 also shows why adding $P(C, Z_m)$ or every model–model pair does not repair an outcome-paired release. For five binary candidates, the separate candidate triples contain 40 raw cells versus 128 for the full joint; for 20 candidates the counts are

160 versus $2^{22} = 4{,}194{,}304$. These are storage dimensions, not privacy guarantees. The executable witness reports value gaps 0.693147 (log), 0.25 (Brier), and 0.292893 (spherical), with zero error in every proper-subset marginal.

For completeness, define the dense-cell cost of separately serializing scope S as $\prod_{X \in S} d_X$, where every alphabet size is an integer at least two. Consider any new scope containing Y, C and a nonempty set T of as-yet unidentified candidates. Extra variables can only increase its cost, and

$$d_Y d_C \prod_{m \in T} d_m \;\geq\; d_Y d_C \sum_{m \in T} d_m.$$

For $|T| = 2$ this follows from $(d_i - 1)(d_j - 1) \geq 1$; the general case follows by induction. Summing over any collection whose covered candidate sets span all missing candidates gives the lower bound $d_Y d_C \sum_m d_m$ (duplicate coverage only increases the sum). One separate (Y, C, Z_m) table per missing candidate attains it. Equality is not generally unique; in particular, one scope containing two binary candidates costs the same 16 cells as their two binary triples. The executable planner returns an attaining repair under exactly this accounting contract. Thus the repair is candidate-linear under dense-cell accounting only: this is linear in candidate count, with cost $\sum_m |\mathcal{Y}||\mathcal{C}||\mathcal{Z}_m|$, not necessarily in the serialized size of a complete joint table.

The accompanying implementation audit finds zero discrepancies on 32,768 raw-marginal schemas, 65,536 identification decisions, and 7,776 grouped-table repair problems (five candidates; alphabet sizes $\{2, 3, 5\}$); these finite checks supplement, rather than replace, the scope and dense-cell proofs (`scripts/audit_release_schema_theorem.py`).

D.8 Exact factorization for a finite count artifact

Let a release contain integer count tables N_{YC} and N_{YZ_m} with the same outcome totals and common sample size n. For candidate m, let $\mathcal{A}_m^{(n)}$ be the finite set of conditional values attained by all nonnegative integer triple tables N_{YCZ_m} having those two margins.

Proposition 5 (Same-size integer factorization) *The exact value-vector set over all n-row datasets preserving the released count tables is*

$$\mathcal{A}_1^{(n)} \times \cdots \times \mathcal{A}_M^{(n)}.$$

Thus selecting one integer-attainable value for each candidate never requires fractional observations or an expanded denominator.

Proof. Projection of any compatible row-level dataset gives one feasible integer triple per candidate, proving one inclusion. Conversely, select an integer triple table for every candidate. Within each (y, c) block of $N_{YC}(y, c)$ rows, candidate m only prescribes a multiset containing the declared number of entries for each z_m symbol. Assign that multiset in any order within the block, independently for every candidate column. All columns coexist on the same rows; the resulting dataset has exactly n rows and recovers every selected triple, hence every released count table. This proves the reverse inclusion. □

This proposition is an exact statement about the observed finite artifact. It does not provide confidence coverage for the population law that may have generated that artifact.

Define the finite-artifact signed margin $\gamma_{n,\mathrm{worst}}$ by replacing each interval $\mathcal{I}_m$ above with its integer-attainable set $\mathcal{A}_m^{(n)}$, and let $\rho_n = \max\{0, \gamma_{n,\mathrm{worst}}\}$. For a fixed permutation and queried gap g, process candidates from worst to best, choosing first the smallest attainable value and then the smallest value at least g above the previous choice.

Proposition 6 (Exact discrete gap feasibility) *The fixed-permutation greedy recursion succeeds if and only if the supplied finite attainable sets contain a value vector in that order with every adjacent gap at least g. Consequently, enumerating permutations and searching g yields the permutation margin and hence $\gamma_{n,\mathrm{worst}}$ and ρ_n; every feasible positive lower endpoint is accompanied by an exact n-row witness through same-size integer factorization.*

Proof. The greedy value at the lowest rank is no larger than the corresponding value in any feasible vector. Inductively, if its value at one rank is no larger, the smallest attainable value at least g above it

is no larger than the value used by any feasible vector at the next rank. Thus greedy failure at any rank precludes every feasible continuation; success directly supplies a witness. Proposition 5 glues the selected integer triples on the original rows. □

Our implementation bisects g and reports both a feasible lower witness and a numerical upper bracket. Row and released-count compatibility are exact; the reported score-valued radius bracket is numerical.

D.9 Proof of Theorem 11

Proof. Fix a best-to-worst permutation $\pi = (\pi_1, \dots, \pi_M)$ and let

$$g_\pi = \min_{1 \le i < j \le M} \frac{U_{\pi_i} - L_{\pi_j}}{j - i}.$$

If a compatible vector has every adjacent gap at least g, telescoping from rank i to rank j gives $v_{\pi_i} - v_{\pi_j} \ge (j-i)g$. Since $v_{\pi_i} \le U_{\pi_i}$ and $v_{\pi_j} \ge L_{\pi_j}$, necessarily $g \le g_\pi$.

Conversely, set the worst value to $v_{\pi_M} = L_{\pi_M}$ and proceed upward:

$$v_{\pi_i} = \max\{L_{\pi_i}, v_{\pi_{i+1}} + g_\pi\}, \qquad i = M-1, \dots, 1.$$

Unrolling the recursion gives

$$v_{\pi_i} = \max_{j \ge i}\{L_{\pi_j} + (j-i)g_\pi\}.$$

The $j = i$ term is at most U_{π_i}. Every $j > i$ term is also at most U_{π_i} by the definition of g_π. Thus the greedy vector lies in $\prod_m [L_m, U_m]$ and has every adjacent gap at least g_π. Theorem 9 realizes it, proving $\gamma_\pi = g_\pi$.

Let $\Delta = \min_{a \ne b}(U_a - L_b)$. If $\Delta \ge 0$, then for every permutation and every $i < j$,

$$\frac{U_{\pi_i} - L_{\pi_j}}{j - i} \ge \frac{\Delta}{M - 1}.$$

Place a pair (a, b) attaining Δ first and last to achieve equality, so $\gamma_{\text{worst}} = \Delta/(M-1)$. If $\Delta < 0$, division by any $j - i \ge 1$ makes a negative numerator weakly less negative, so every normalized cross difference is at least Δ. Placing an attaining pair adjacently achieves equality, and $\gamma_{\text{worst}} = \Delta$. In both cases $\rho = \max\{0, \gamma_{\text{worst}}\}$.

For $\epsilon \ge 0$, all permutations support adjacent gaps strictly greater than ϵ exactly when $\gamma_{\text{worst}} > \epsilon$ (equivalently $\rho > \epsilon$ for $\epsilon \ge 0$), equivalently $\Delta > (M-1)\epsilon$. Finally, if all intervals share $[L, U]$, then $U_a - L_b \ge U - L$ for every $a \ne b$; hence the earlier common-overlap certificate is an immediate, possibly strict, corollary. □

Executable verification summary. Table 4 separates mathematical obligations from finite checks. The complete per-world ledgers, symbolic residuals, exact counts, and solver states remain in the author-side audit archive; the table records the quantities needed to interpret the claims.

Table 4: Compact audit of the constructions. Exact rows means exact reconstruction of released empirical counts, not population inference.

Audit	Result	What it establishes
Calibrated forecast lift	360 five-candidate worlds; 120/120 orders for each of log, Brier, and spherical loss; residuals at most 2.22×10^{-16}	Finite implementation check of the analytic smoothing construction
Complete-joint / local tangent	120/120 orders at five shrinking radii; centered Jacobian rank four; maximum release error 1.11×10^{-16}	Checks the fixed-release quantifiers and first-order scaling
Exact algebra / symbolic identities	Rational three-candidate release invariance and TV identity; eleven symbolic residuals reduce exactly to zero	Independent algebraic cross-check, not a computer proof
ImageNet-16H (I16-JOINT)	Log: 120/120 exact-row orders; Brier: 119/120; spherical: 117/120 at $n = 4{,}716$	Post-hoc finite-artifact support; no cross-score prevalence claim
CIFAR-10H (C10-JOINT)	Log: 24/24 exact-row orders at $n = 10{,}000$; every candidate can win within TV radius 0.0056, and all orders by 0.0102	Locality around the observed completion; counts are lower bounds until the full order family is reached

The synthetic calculations test implementation identities, while global smoothness, interior overlap, calibration, and nondecodability remain analytic proof obligations. The real-artifact rows reuse opened data and therefore do not estimate prevalence, causal plausibility, or population probability.

Proposition 7 (Exact distance to material fixed-rule regret) *On the continuous fiber $\mathcal{C}(\widetilde{\mathcal{O}})$, fix an aligned reference Q_0, linear fixed-rule risks $R_m(Q) = \langle r_m, Q\rangle$, a selection a, and $\epsilon > 0$. The minimum distance to ϵ simple regret is*

$$\rho_\epsilon(Q_0, a) = \min_{j \neq a} \min_{\substack{Q \in \mathcal{C}(\widetilde{\mathcal{O}}) \\ \langle r_a - r_j, Q\rangle \geq \epsilon}} \tfrac{1}{2}\|Q - Q_0\|_1.$$

D.10 Proof and certificate for Proposition 7

For any completion Q and fixed selection a,

$$R_a(Q) - \min_j R_j(Q) = \max_j \langle r_a - r_j, Q\rangle = \max\left\{0, \max_{j \neq a} \langle r_a - r_j, Q\rangle\right\}.$$

Therefore the adverse set at threshold ϵ is the finite union over $j \neq a$ of the release-fiber halfspaces $\{Q : \langle r_a - r_j, Q\rangle \geq \epsilon\}$. Distance to a finite union is the minimum of the component distances, which gives the displayed formula for $\rho_\epsilon(Q_0, a)$.

Write the release constraints as $AQ = b$. For a fixed challenger, introduce nonnegative auxiliaries u with $Q - Q_0 \leq u$ and $Q_0 - Q \leq u$. Minimizing $\frac{1}{2}\mathbf{1}^\top u$ subject to $AQ = b$, $Q \geq 0$, and the challenger halfspace is exactly the stated TV problem. It is a finite LP; infeasibility means that challenger cannot produce ϵ regret anywhere in the release fiber. A primal optimum supplies the nearest compatible witness, and the standard LP dual supplies a matching lower certificate.

For $d_j = r_a - r_j$, define $\operatorname{osc}(d_j) = \max_x d_j(x) - \min_x d_j(x)$. Both Q and Q_0 have unit mass, so the total-variation oscillation inequality gives

$$\langle d_j, Q - Q_0\rangle \leq \operatorname{osc}(d_j)\, \mathrm{TV}(Q, Q_0).$$

If the reference contrast is below ϵ, any feasible adverse completion must increase it by at least $[\epsilon - \langle d_j, Q_0\rangle]_+$; if it already reaches ϵ, the zero lower bound is immediate. Combining these inequalities gives the corresponding TV lower bound. If $\operatorname{osc}(d_j) = 0$ and the numerator is positive, the halfspace is infeasible. This proof uses only fixed finite loss coefficients. For Bayes value or a refitted rule, the risk map is generally nonlinear and the LP identity does not follow.

The reference Q_0 is not recoverable from a deficient release in general. It must be supplied by an aligned evaluation, or stated explicitly as the center of an assumption-indexed sensitivity analysis. The certificate therefore answers how far the declared decision state is from material failure rather than selecting a privileged completion from the release fiber.

Corollary 6 (Operational consequences of rank shattering) *Let $\{Q_\pi : \pi \in S_M\}$ expose one common release, with Q_π having strict risk order π, and define the common winner gap*

$$\gamma := \min_{\pi \in S_M} \min_{a \neq \pi(1)} \{R_a(Q_\pi) - R_{\pi(1)}(Q_\pi)\} > 0.$$

On this family, randomized minimax exact-ranking error is $1 - 1/M!$ and normalized Kendall error is $1/2$. Every release-only forced winner selector has a selector-dependent world with selection error at least $1 - 1/M$ and expected simple regret at least $\gamma(1 - 1/M)$. Consequently, for every $0 \leq \epsilon < \gamma$, that world also has probability at least $1 - 1/M$ of simple regret strictly greater than ϵ.

D.11 Proof of Corollary 6

For each strict permutation choose one positive-gap witness supplied by the shattering result, and restrict attention to this $M!$-world subfamily. Every selected world exposes the same $\mathcal{O}$, so any possibly randomized release-only rule has the same output distribution in all selected worlds.

For exact ranking, put a uniform prior on the $M!$ compatible permutations. The average probability of matching the true permutation is

$$\frac{1}{M!} \sum_{\pi} P(\widehat{\pi} = \pi) = \frac{1}{M!}.$$

Some selected compatible world therefore has error at least $1 - 1/M!$. Uniform random ranking attains that error against every strict permutation, proving the randomized minimax equality on the selected subfamily. A deterministic rule is wrong in at least one selected world. On a larger compatible class this remains a lower bound; if that class contains ties, claiming equality would additionally require a convention for exact-ranking loss on tied truths.

For Kendall error, every unordered pair is reversed with probability one half under a uniformly random true permutation, regardless of any fixed predicted order. Averaging additionally over a randomized prediction preserves this value. The average Kendall distance is therefore $\binom{M}{2}/2$; after normalization by $\binom{M}{2}$, every rule has worst-world error at least $1/2$. A uniformly random predicted ranking has expected normalized distance exactly $1/2$ against every true permutation, proving equality.

For winner selection, retain M worlds in which each candidate is the unique winner. Because all worlds expose the same release, a possibly randomized selector uses one common action distribution $(p_1, \ldots, p_M)$. Choose m with $p_m \leq 1/M$ and a retained world whose winner is m. Its selection error is at least $1 - 1/M$, and every incorrect action incurs regret at least γ. The expected simple regret is therefore at least $\gamma(1 - 1/M)$. If $0 \leq \epsilon < \gamma$, every incorrect action also lies in the strict event $\{\text{regret} > \epsilon\}$, proving the probability claim. At $\epsilon = \gamma$ the same argument only guarantees $\text{regret} \geq \gamma$. Abstention avoids a false direction but is excluded from this forced-selection decision class.

Finally, let a benchmark reveal only a deterministic summary $T(\mathcal{O})$, such as context and candidate marginal risks or their induced rank. All constructed worlds share the same $\mathcal{O}$ and therefore the same $T(\mathcal{O})$, so every preceding indistinguishability and decision argument continues unchanged. This monotonicity does not cover a summary computed from aligned (Y, C, Z_m) entities, which can encode the missing coupling.

D.12 Finite-artifact evidence and claim boundary

For binary theorem-matched artifacts, Fréchet bounds reduce each compatible triple to a rectangle, so boundary and stationary-point enumeration gives exact candidate-wise projected intervals under the weaker release. Glued witnesses under that contract realize 120/120 integer orders on ImageNet-16H (I16-MARG, $n = 4{,}716$) and 24/24 on CIFAR-10H (C10-MARG, $n = 10{,}000$), with every released count reconstructed exactly. These are finite-artifact compatibility statements, not

sampling or prevalence conclusions, and do not establish feasibility with the complete candidate law fixed. The separate complete-joint CIFAR ledger (C10-JOINT) also verifies 24/24 orders; the author-side manifest indexes it as `cifar10h_joint_prediction_release_v1.json`. For the submitted archive, `evidence/witness_index.json` maps the I16-MARG, C10-MARG, and C10-JOINT identifiers to their public evidence records and integrity hashes.

Native probabilities and fixed rules. To retain confidence geometry, a separate ImageNet-16H audit fixes $Q_m(k \mid p_m, c) \propto p_m(k)\widehat{P}(C = c \mid Y = k)$, with Jeffreys smoothing computed from released $P(Y, C)$. Candidate selection is also release-only: rank all models by native marginal log loss, retain the best half, and choose the four consecutive ranks with the smallest log-loss span.

Table 5: Finite-artifact evidence retained in the compact appendix. Counts refer to exact-row witnesses unless stated otherwise.

Artifact	Result	Boundary
ImageNet native probabilities, clean (I16-NATIVE)	24/24 orders for each of log, Brier, and spherical loss; centered log-risk simplex inradius exceeds 0.0069 nats	Post-hoc fixed-rule discovery; serialized-to-intended risk transfer is interval certified
ImageNet native probabilities, frozen shift (I16-SHIFT)	2/24 positive exact-row witnesses under each score, with one shared winner	Without dual or integer infeasibility certificates, we do not interpret $2/24$ as an exact capacity theorem
CIFAR native-ten stronger release	First requested order fails; all eight SLSQP starts have non-success status	Frozen checkpoint, fail-closed reconstruction; neither a positive witness nor an infeasibility certificate
ACIV CXR, score-uniform Blackwell	120/120 exact orders, 960 strictly interior rational kernels; smallest log gap 0.0374323 nats and kernel entry $7/48416$	Post-hoc reuse; exact compatibility, not local or clinical evidence
UCI Adult, result-blind protocol	120/120 continuous and integer orders; smallest exact-row gap 0.0026417	Repository/commit-frozen finite-artifact evidence, not a fairness or population claim
Observed-alignment panel	Skin supplies two certified pair reversals; ImageNet/CIFAR confirmations are preserved or unresolved	The panel was not population-sampled and does not estimate prevalence in either direction

For the clean ImageNet log release, four retained risk vertices admit positive tie weights and centered rank three; a 60-digit interval audit transfers the serialized inradius to an intended-risk lower bound 0.006899999998. Conversely, the Covertype release has full centered rank but an exact decimal-rational dual certificate gives compatible value span 0.0285074412, so it has no compatible common tie. This makes the tie and rank gates empirically separable.

The same byte-frozen 4,800-row ImageNet log release has a decision-level consequence. Loading the author-side 24 exact-row assignments identified by their hashes and using 60-digit outward loss intervals gives an adjacent-gap lower bound 0.0142553320 nats; direct witnesses already imply $\mathfrak{G}(\Gamma) > 0.0805400041$ nats. Theorem 3's finite-quota top-k calculation over all compatible row assignments sharpens this to $0.0939264361 < \mathfrak{G}(\Gamma) < 0.0939264362$ nats. The audit checks the forecast bytes, release counts, and risk orderings, with an independent floating-point top-k implementation agreeing to 10^{-9}. Consequently no rule among these four is release-certified at $\epsilon = .005$. This post-hoc, worst-compatible result is not a realized wrong choice, query-cost estimate, or claim about the shifted release.

Menu expansion does not erase this deficiency. On the same release, 60-digit interval audits give deterministic regret $0.0246202164 < \mathfrak{G}(\Gamma) < 0.0246202165$ for 20 fused models plus a context-only rule, and $0.0502612817 < \mathfrak{G}(\Gamma) < 0.0502612818$ after adding 20 normalized un-

fused forecasts. A larger comparator set can raise regret. Two compatible worlds per menu with rational mixing weights also lower-bound every randomized action on the 4/21/41-rule menus by 0.0489242142/0.0204100113/0.0295673047 nats. The 60-digit interval certificate serializes both assignments; these are lower bounds, not exact mixed minimax values or claims about arbitrary fusion. All five menus select the actual aligned winner (zero observed regret): certificate deficiency, not choice repair.

Exact score-uniform certificate. The public source is Guo et al.'s supplementary `CXRForecasting.csv` Guo et al. (2026). Its 12,229 unique study IDs define a binary abnormality task: $Y = \texttt{abnormal_stratified}$, C is radiologist `diagnosis` probability thresholded at 0.5, and $Z_{1:5}$ are the 0.5-thresholded outputs of DenseNet-121, Inception-v3, Swin-B, ResNet-50, and ViT-B/16. The frozen release inventory is

$$N_{Y,C} = \begin{pmatrix} 6809 & 1492 \\ 1884 & 2044 \end{pmatrix}, \qquad (N_{Z_1=1}, \ldots, N_{Z_5=1}) = (3665, 3372, 3377, 3649, 3527),$$

with rows $Y = 0, 1$ and columns $C = 0, 1$; the hashed manifest stores the complete 2×32 table $N_{Y,Z_{1:5}}$. Thus the hidden object is only the pairing of C with the complete candidate vector within outcome strata.

One independently replayed adjacent-pair witness, DenseNet-121 $\succ_B$ Inception-v3, uses the exact context-specific kernels

$$K_0 = \begin{pmatrix} 2326940/2338927 & 11987/2338927 \\ 1188723/11694635 & 10505912/11694635 \end{pmatrix},$$

$$K_1 = \begin{pmatrix} 485369/488062 & 2693/488062 \\ 38903/488062 & 449159/488062 \end{pmatrix}.$$

Both are strictly interior row-stochastic matrices; right multiplication maps the DenseNet channel to the Inception channel exactly in each context. The 80-digit lower bound on this adjacent log-value gap is $0.06335507916706089712406603404$. For every CXR order, the artifact supplies two context-specific rational garbling kernels for each adjacent pair. A non-importing verifier reconstructs the released count table, checks all stochastic-kernel identities, recomputes log gaps at 80-digit precision, and lifts the counts to the 12,229 study IDs. The included release manifest identifies the full 120-order ledger and independent checker. The earlier 119/120 run and the deterministic certificate-selection correction are both retained.

Result-blind Adult audit. Sex and census weight are withheld from candidate inputs. The protocol fixes the official split, five estimators, log score, starts, and gates before target couplings are computed. An independent contract test reconstructs the released 2-by-32 table and all continuous and integer values. The included release manifest records the training, prediction, and verification receipts.

Release surfaces and traceability. Official ranking surfaces for WILDS, RobustBench, CheXpert, ImageNet ILSVRC, and Grand Challenge expose aggregate metrics rather than a common aligned $(Y, C, Z_{1:M})$ table. Some ecosystems permit checkpoint reruns or new aligned audits; the claim is only that the ranking surface itself cannot answer the context-assisted question. A trusted evaluator can repair the release by constructing aligned outcomes, context, and candidate outputs.

Author-side records retain, for iWildCam, the formal summary, audit, certificate, and fail-closed checker and, for MOLHIV, the primary certificate, audit, and serialization checker. These are traceability entries, not additional positive experiments. Full optimizer histories, negative sensitivities, exact matrices, row assignments, and correction ledgers remain machine-readable in the artifact rather than being repeated in the PDF.

E FINITE-ALPHABET AND STABLE-REFIT EXTENSIONS

This optional extension treats finite-alphabet fixed rules, locally stable population refits, common strongly convex heads, and split-sample persistence. Each states its own stability contract; these extensions are not needed to verify the frozen-rule disclosure certificates.

E.1 FINITE-ALPHABET FIXED RULES

Proposition 8 (Finite-alphabet fixed-rule criterion) *Let $\mathcal{C}$ and $\mathcal{Z}$ be nonempty finite sets, and let one possibly randomized fixed rule have finite real-valued expected loss $h(y,c,z) = \mathbb{E}_U[\ell\{y, \mathcal{A}(c,z;U)\}]$. The rule's risk agrees under every two full laws having the same released pair margins $P(Y,C)$ and $P(Y,Z)$, for every possible compatible pair of margins, if and only if, for every y, c, c', z, z',*

$$\Delta_y(c,c';z,z') = h(y,c,z) - h(y,c,z') - h(y,c',z) + h(y,c',z') = 0.$$

Equivalently, for each y there are functions a_y, b_y such that $h(y,c,z) = a_y(c) + b_y(z)$. If one mixed difference Δ is nonzero, then for every finite $M \geq 2$ there exist one *full-support release fixing $P(Y,C)$ and the complete $P(Y,Z_{1:M})$, one compatible completion Q_0, and $\eta_0 > 0$ such that, for every $0 < \eta < \eta_0$ and every permutation, the same rule applied to each Z_m realizes that strict risk order within TV distance η of Q_0. The construction has adjacent gap $|\Delta|\eta/(2A_M)$, where $A_M = \lfloor M^2/4 \rfloor$.*

Proof. If all mixed differences vanish, fixing reference levels c_0, z_0 gives

$$h(y,c,z) = h(y,c,z_0) + h(y,c_0,z) - h(y,c_0,z_0),$$

so the expectation is a functional of the two pair laws. Conversely, any nonzero mixed difference changes along the corresponding 2×2 signed cycle while leaving both margins fixed.

For simultaneity, choose the nonzero cycle at y_0 and define u on $\mathcal{C}$ to be $+1, -1$ at c, c' and zero elsewhere; define v analogously on $\mathcal{Z}$ at z, z'. Start from the full-support law under which, given Y, $C, Z_1, \ldots, Z_M$ are mutually independent and uniform. For sufficiently small $\theta \in \mathbb{R}^M$, set

$$Q_\theta(y,c,z_{1:M}) = Q_0(y,c,z_{1:M}) \left[1 + \mathbf{1}\{y = y_0\} \sum_{m=1}^{M} \theta_m u(c) v(z_m)\right].$$

Because u and v each sum to zero, summing over C removes every perturbation and summing over $Z_{1:M}$ does the same. Terms indexed by $j \neq m$ vanish from (Y, C, Z_m). Hence

$$R_m(\theta) = R_0 + \frac{P(Y = y_0)}{|\mathcal{C}||\mathcal{Z}|} \Delta\, \theta_m.$$

The triangle inequality gives $\mathrm{TV}(Q_\theta, Q_0) \leq 2P(Y = y_0)\|\theta\|_1/(|\mathcal{C}||\mathcal{Z}|)$. Assign centered, equally spaced coordinates in the requested increasing-risk order, oriented by $\mathrm{sgn}(\Delta)$, with step $\eta|\mathcal{C}||\mathcal{Z}|/[2P(Y = y_0)A_M]$. This proves the TV bound and the stated gap. For example, $\eta < 2P(Y = y_0)/(|\mathcal{C}||\mathcal{Z}|)$ makes every perturbation bracket strictly positive, so shrinking η preserves full support. □

E.2 STABLE POPULATION REFITS

The next result separates what can be proved for refitting from the false claim that every training procedure must shatter. Let $P_t = P_0 + tD$ be the candidate-wise fixed-margin path induced by the cycle above. Let $h_\beta(y,c,z) = \ell\{y, a_\beta(c,z)\}$ and suppose a population training rule returns

$$\widehat{\beta}(t) = \arg\min_\beta J(\beta, t), \qquad J(\beta,t) = \mathbb{E}_{P_t} h_\beta + \Omega(\beta).$$

Proposition 9 (Locally stable population refitting) *Suppose the parameter space is open near β_0, h_β and Ω are twice continuously differentiable there, $\widehat{\beta}(0) = \beta_0$ is a unique interior global optimum, and $H = \nabla^2_\beta J(\beta_0, 0)$ is positive definite. Assume in addition that the population training rule is* optimizer-stable *along the path: for all sufficiently small $|t|$, a selected global minimizer $\widehat{\beta}(t)$ exists and $\widehat{\beta}(t) \to \beta_0$ as $t \to 0$. Define the unpenalized fitted evaluation risk $r(t) = \mathbb{E}_{P_t} h_{\widehat{\beta}(t)}$. Then r is continuously differentiable locally and*

$$r'(0) = \langle D, h_{\beta_0} \rangle - \{\mathbb{E}_{P_0} \nabla_\beta h_{\beta_0}\}^\top H^{-1} \langle D, \nabla_\beta h_{\beta_0} \rangle. \tag{19}$$

If this auditable end-to-end derivative is nonzero, then for every finite M the same refitted algorithm realizes all $M!$ strict evaluation-risk orders in an arbitrarily small common-release neighborhood.

Proof. The implicit-function theorem applied to $F(\beta, t) = \nabla_\beta J(\beta, t)$ gives a unique continuously differentiable critical branch $\beta_*(t)$ in a neighborhood of $(\beta_0, 0)$, with

$$\beta_*'(0) = -H^{-1}\langle D, \nabla_\beta h_{\beta_0}\rangle.$$

Optimizer stability places $\widehat{\beta}(t)$ in this neighborhood for all sufficiently small $|t|$. An interior global minimizer satisfies the first-order condition, so local uniqueness of the critical branch gives $\widehat{\beta}(t) = \beta_*(t)$. Thus $\widehat{\beta}$ is continuously differentiable locally. Differentiating the unpenalized evaluation risk yields (19). If its derivative is nonzero, continuity makes r strictly monotone on some interval around zero. In the simultaneous construction of Proposition 8, candidate m has pair law P_{θ_m} and therefore fitted risk $r(\theta_m)$, while the released $P(Y, C)$ and complete $P(Y, Z_{1:M})$ remain unchanged. Assigning distinct coordinates in any requested order inside the monotone interval realizes that order; scaling them toward zero makes the common TV radius arbitrary. □

When $\Omega = 0$ and the training and evaluation losses coincide, the second term in Equation (19) vanishes by first-order optimality and the condition reduces to a mixed loss interaction at the fitted rule. With regularization, Equation (19) also accounts for optimizer response. The executable audit uses a three-level context, four-level signal, and an L_2-regularized additive logistic probe. It reoptimizes each distinct pair law, verifies a positive-definite Hessian and the analytic derivative against finite differences, and realizes 120/120 orders for five candidates with maximum release error below 6×10^{-17}. The conclusion is local and applies to the stated optimizer-stable population training rule, not arbitrary nonconvex or discontinuous training procedures. Uniform strong convexity and coercivity of $J(\cdot, t)$ for all sufficiently small $|t|$ are one checkable sufficient route to the optimizer-stability assumption; the proposition does not claim they hold automatically.

Corollary 7 (Global strongly-convex population refitting) *In Proposition 9, take the parameter space to be $\mathbb{R}^d$. Suppose that, for all sufficiently small $|t|$, $J(\cdot, t)$ is twice continuously differentiable and μ-strongly convex for one constant $\mu > 0$, and that $\nabla_\beta J(\beta_0, t)$ is continuous at $t = 0$. Then a unique global minimizer $\widehat{\beta}(t)$ exists, optimizer stability is automatic, and*

$$\|\widehat{\beta}(t) - \beta_0\|_2 \le \mu^{-1}\|\nabla_\beta J(\beta_0, t)\|_2 \longrightarrow 0.$$

Consequently the derivative formula and local all-M! conclusion of Proposition 9 hold without separately assuming that the selected optimizer stays on the local IFT branch.

Proof. Strong convexity gives a coercive quadratic lower bound, hence existence and uniqueness of the global minimizer. Strong monotonicity of the gradient and $\nabla J\{\widehat{\beta}(t), t\} = 0$ give

$$\mu\|\widehat{\beta}(t) - \beta_0\|_2^2 \le \left\langle -\nabla J(\beta_0, t), \widehat{\beta}(t) - \beta_0 \right\rangle,$$

which implies the displayed bound. Continuity puts the unique global minimizer inside the IFT neighbourhood, where it equals the unique critical branch used in Proposition 9. □

E.3 Convex heads and analytic genericity

Common frozen-encoder convex heads. Let a frozen encoder and declared context map produce a bounded finite-dimensional feature vector. Binary logistic, multinomial softmax, and ridge heads with a fixed L_2 penalty $\lambda\|\beta\|_2^2/2$, $\lambda > 0$, have a twice continuously differentiable population objective whose Hessian is the convex data-loss Hessian plus λI. They are therefore globally λ-strongly convex; on finite support their gradient varies continuously with the coupling. Corollary 7 consequently applies whenever the total coupling derivative in Proposition 9 is nonzero. This specialization covers a common frozen-representation evaluation stack, but not tuning λ on the target, selecting among heads, fine-tuning the encoder, or nonconvex end-to-end training.

Corollary 8 (Conditional genericity for analytic convex heads) *Fix the finite support, the coupling cycle D, and a connected open parameter set $\Xi \subset \mathbb{R}^p$. Suppose the cell masses, bounded frozen features, training loss, evaluation loss, and fixed regularizer depend real-analytically on $\xi \in \Xi$ and on the finite-dimensional head parameter β. Suppose also that the population training objective is uniformly μ-strongly convex in β, and define $G(\xi)$ to be the total coupling derivative in Equation* (19). *Then G is real analytic. If one declared witness $\xi_\star \in \Xi$ satisfies $G(\xi_\star) \neq 0$, the degenerate set*

$$\{\xi \in \Xi : G(\xi) = 0\}$$

has Lebesgue measure zero. Hence, outside that set, the same fixed refitting algorithm has the local all-$M!$ conclusion of Proposition 9.

Proof. The score equation $\nabla_\beta J(\beta, 0; \xi) = 0$ is real analytic, while uniform strong convexity makes its β-Jacobian invertible and gives a unique global solution. The analytic implicit-function theorem therefore makes $\widehat{\beta}(0; \xi)$ locally analytic in ξ; uniqueness patches these local branches into one analytic map on Ξ. Every term in Equation (19), including the inverse Hessian, is then analytic, so G is analytic. The nonzero witness shows that G is not the zero analytic function. The zero set of a nonzero real-analytic scalar function on a connected open set has Lebesgue measure zero. For every point outside that set, Proposition 9 applies. □

This is a conditional genericity statement inside a declared analytic family, not a claim that a chosen coupling direction interacts with every feature map or that arbitrary training pipelines are generically unstable. The executable audit supplies nonzero-derivative witnesses for L_2-regularized binary logistic, multinomial softmax, and ridge heads and checks each analytic derivative against finite differences.

E.4 Split-sample finite-sample persistence

The next statement separates population nonidentification from sampling noise. It is intentionally a split-sample result: training and evaluation entities are independent, as in the source-fit/held-out-target workflows used in the empirical supplement. Candidates may share evaluation rows and have correlated loss estimates; the union bound below does not require candidate estimates to be independent.

Proposition 10 (Finite-sample persistence for stable refits) *Let $\mathcal{W}$ be a finite collection of population ranking worlds and let r_{wm} be the population fitted evaluation risk of candidate m in world w. Suppose every requested adjacent population gap is at least $g > 0$. For each (w, m), let the population training objective be μ-strongly convex with minimizer β_{wm}, and fit on an independent training sample. Assume that, simultaneously over $K = |\mathcal{W}|M$ fits, an event of probability at least $1 - \delta_{\mathrm{tr}}$ satisfies*

$$\lambda_{\min}\{\nabla^2 J_{n,wm}(\beta)\} \geq \mu/2 \quad \textit{along } [\beta_{wm}, \widehat{\beta}_{wm}], \qquad \|\nabla J_{n,wm}(\beta_{wm})\|_2 \leq a_n.$$

If the evaluation loss is in $[0, B]$ and is L-Lipschitz in β, then an independent evaluation sample of size N ranks every world correctly with probability at least $1 - \delta_{\mathrm{tr}} - \delta_{\mathrm{ev}}$ whenever

$$2e_{n,N} < g, \qquad e_{n,N} = \frac{2La_n}{\mu} + B\sqrt{\frac{\log(2K/\delta_{\mathrm{ev}})}{2N}}.$$

For the $M!$ witnesses of Proposition 9, this is a simultaneous high-probability persistence statement with $K = M\,M!$; it is not a claim about arbitrary SGD trajectories or nonconvex representation learning. The inequality $2e_{n,N} < g$ says that the combined training and evaluation error cannot erase the smallest constructed adjacent ranking gap.

Proof. On the training event, strong monotonicity of the empirical gradient and the empirical first-order condition imply $\|\widehat{\beta}_{wm} - \beta_{wm}\|_2 \leq 2a_n/\mu$. Lipschitzness therefore bounds the population evaluation-risk displacement by $2La_n/\mu$. Conditional on the fitted parameters, Hoeffding's inequality and a union bound over the K held-out risk estimates bounds their evaluation-sampling errors by the second term in $e_{n,N}$ with probability at least $1 - \delta_{\mathrm{ev}}$. Thus every reported risk is within $e_{n,N}$ of its population counterpart. Any adjacent difference moves by at most $2e_{n,N} < g$, so all requested strict orders persist simultaneously. □

An approximate empirical optimizer with objective suboptimality at most $\varepsilon_{\mathrm{opt}}$ adds at most $2L\sqrt{\varepsilon_{\mathrm{opt}}/\mu}$ to $e_{n,N}$ under the same empirical curvature event. Finite source-only hyperparameter families can be covered by including every frozen fit in K; data-dependent architecture search and end-to-end nonconvex training remain outside the theorem.

F ONLINE DISCOVERY OF A CERTIFICATE

This optional planning extension starts where the holder-known certificate problem ends. A holder's shortest proof assumes the true alignment is known. The online evaluator instead chooses queries from public information and answers seen so far, using the same deterministic stopping verifier.

F.1 EXACT ONLINE REPAIR ON PUBLIC TYPE COUNTS

Prior-free minimax quotient. For a finite binary-context fixed-rule release, define $i \sim i'$ when rows i, i' have the same outcome and the same complete candidate-by-context loss vector. These classes use only released information. Let n_t and y_t be a type's size and outcome, and let K_y be the released number of context-one rows in outcome stratum y. For state $s = (s_{t1}, s_{t0})_{t=1}^{T}$, the feasible total-one counts are

$$\mathcal{K}(s) = \left\{ k \in \mathbb{Z}^T : s_{t1} \le k_t \le n_t - s_{t0}, \quad \sum_{t:y_t=y} k_t = K_y \ \forall y \right\}.$$

Every compatible labeled completion projects to one $k \in \mathcal{K}(s)$, and every such k has a labeled realization. Its additive risk is

$$R_a(k) = \frac{1}{N} \sum_t \{k_t \ell_a(t,1) + (n_t - k_t)\ell_a(t,0)\}.$$

Let $\Gamma(s)$ be the corresponding completion fiber, set

$$G(s) = \min_a \max_j \sup_{c \in \Gamma(s)} \{R_a(c) - R_j(c)\},$$

and let $\mathcal{F}_t(s)$ contain the answers leaving $\Gamma(s + e_{tc})$ nonempty. The sharp top-k audit computes $G(s)$ without enumerating completions.

Theorem 12 (Exact public-type quotient for online repair) *With the notation above, the optimal worst-case number of further row queries is*

$$D(s) = \begin{cases} 0, & G(s) \le \epsilon, \\ 1 + \min\limits_{t: s_{t1}+s_{t0}<n_t} \max\limits_{c \in \mathcal{F}_t(s)} D(s + e_{tc}), & G(s) > \epsilon. \end{cases}$$

Querying any hidden representative of a minimizing type is row-level minimax optimal, and the quotient contains at most $\prod_t (n_t + 1)(n_t + 2)/2$ *states.*

Proof of Theorem 12. Permuting row identities within a public type gives a risk-preserving bijection between completion fibers and conjugates any row-ID-aware continuation policy to the permuted transcript. Thus two transcripts with the same s have isomorphic minimax subproblems. If $G(s) \le \epsilon$, the sharp audit supplies a common ϵ-optimal action. Otherwise any row-level policy must query a row of some type t, and an adversary may return any $c \in \mathcal{F}_t(s)$, leaving exactly state $s + e_{tc}$. This gives the Bellman lower bound. Querying any hidden representative of a minimizing type and recursing attains it. Backward induction on the number of hidden rows proves equality and shows that row labels cannot improve minimax depth. Type t has $(n_t + 1)(n_t + 2)/2$ nonnegative count pairs with sum at most n_t; multiplying gives the state bound before margin feasibility removes additional states. □

Prior-dependent HEC quotient. The same quotient also makes a Bayesian decision-region policy exact without listing labeled alignments. Let $\pi(k) > 0$ be a rational full-support prior on the quota-compatible count vectors, and, conditional on k, put the uniform prior on its $\prod_t \binom{n_t}{k_t}$ labeled realizations. For a count state s, define the unnormalized posterior weight

$$W_s(k) = \pi(k) \prod_t \frac{\binom{n_t - s_{t1} - s_{t0}}{k_t - s_{t1}}}{\binom{n_t}{k_t}},$$

with an infeasible binomial coefficient interpreted as zero, and let $A_\epsilon(k) = \{a : R_a(k) - \min_j R_j(k) \le \epsilon\}$.

Theorem 13 (Exact public-type quotient for Bayesian HEC) *Group count vectors having the same nonempty set $A_\epsilon(k)$. At every transcript, the mass of each resulting decision subregion is exactly the sum of $W_s(k)$ over its group. Consequently all Hyperedge Cutting (HEC) weights, one-step expected residual weights, stopping decisions, and query choices can be computed from (s, W_s) and agree with explicit HEC on the full labeled alignment space, up to fixed tie breaking. Querying any hidden representative of a selected type implements the resulting row-level policy.*

Let G be the number of distinct action-set signatures,

$$d = 1 + \min\left\{ \max_g |A_\epsilon(g)|, \max_a |\{g : a \in A_\epsilon(g)\}| \right\}, \qquad p_{\min} = \min_k \frac{\pi(k)}{\prod_t \binom{n_t}{k_t}}.$$

The policy is zero-error on every compatible alignment and inherits the HEC expected-cost bound

$$C(\pi_{\mathrm{HEC}}) \le \{d \log(1/p_{\min}) + 1\}\, C(\pi^*)$$

from Javdani et al. (2014). It uses at most $L \le \prod_t (n_t + 1)$ count vectors, $S \le \prod_t \binom{n_t+2}{2}$ count states, and $H \le \binom{G+d-1}{d}$ subregion hyperedges. Thus for fixed T and M and rationally encoded losses, tolerance, and prior, the quotient computation is polynomial in N and the input bit length, without enumerating the $\sum_k \prod_t \binom{n_t}{k_t}$ labeled alignments.

Proof. Within type t, the likelihood of the observed labeled answers under count k_t is $\binom{n_t - s_{t1} - s_{t0}}{k_t - s_{t1}} / \binom{n_t}{k_t}$; multiplying across types and by $\pi(k)$ gives $W_s(k)$. Risks and therefore decision-region membership depend only on k, so summing these weights by action-set signature gives exactly the subregion masses that explicit HEC obtains by summing over labeled hypotheses. A hidden row in type t has only the two successor count states $s + e_{t0}$ and $s + e_{t1}$; their total weights give the exact branch probabilities. Hence the HEC objective, its greedy maximizer, and its terminal test are functions of the count state alone. Induction over remaining rows proves equality with the explicit policy under common tie breaking and proves zero-error termination because full support leaves no compatible alignment unrepresented. The displayed expected-cost inequality is then Theorem 3 of Javdani et al. (2014) applied to this finite labeled hypothesis space. The three size bounds count, respectively, type-one allocations, feasible observed count pairs, and multisets of d signatures; quota constraints only remove objects from these upper bounds. □

This specializes decision-region querying to fixed outcome–context margins, shared candidate losses, and overlapping ϵ-optimal action regions. Its expected-cost guarantee assumes the stated prior: the factor can be loose when $p_{\min}$ is small and is neither a minimax nor a holder-proof ratio. The state and signature counts S, H can grow exponentially with T or M; polynomial dependence on N for fixed parameters is not necessarily polynomial in succinctly encoded capacities. Thus this is a structured guarantee, not a general scalable planner.

Tolerance-safe coarsening.

Corollary 9 (Tolerance-safe coarsened public-type planning) *Before any answer is observed, let a release-measurable map $\tau(i)$ coarsen the exact public-type partition and merge rows only within the same outcome stratum, and let the approximate loss $\widetilde{\ell}_a(\tau(i), c)$ be constant on each resulting type. Keep the original completion space, row queries, and unit query costs, and suppose*

$$\max_{i,a,c} |\ell_a(i, c) - \widetilde{\ell}_a(\tau(i), c)| \le \delta.$$

Put $\eta = 2\delta$. The Bellman recursion of Theorem 12, applied to $\widetilde{\ell}$, is the exact row-level minimax quotient for the approximate problem and has at most $\prod_t \binom{n_t+2}{2}$ states. If $\epsilon \ge \eta$, its policy at tolerance $\epsilon - \eta$ is zero-error ϵ-safe for the original loss table. If $\epsilon \ge 2\eta$, then at every feasible transcript h, with coarsened count state $s(h)$,

$$D^*_\ell(\epsilon; h) \le D^*_{\widetilde{\ell}}(\epsilon - \eta; s(h)) \le D^*_\ell(\epsilon - 2\eta; h).$$

Every genuine merge of two nonempty exact types strictly decreases this unpruned state bound. If losses lie in $[0, B]$, group rows by outcome and their complete loss vectors after coordinatewise nearest rounding to the grid $\{2q\delta : q = 0, \ldots, \lceil B/(2\delta) \rceil\}$ for $\delta > 0$, with fixed public tie breaking. This yields error at most δ and at most $H(\lceil B/(2\delta) \rceil + 1)^{2M}$ coarsened types across H outcome strata.

Proof. For every completion and candidate, averaging the rowwise bound gives $|R_a - \widetilde{R}_a| \le \delta$. Hence every pairwise risk difference, each candidate's worst compatible regret, and their minimum change by at most $2\delta = \eta$. Permuting row identities within a coarsened type preserves the approximate risks and the quota fiber, so the conjugacy and backward-induction proof of Theorem 12 applies verbatim to the approximate problem. An approximate terminal action at $\epsilon - \eta$ is therefore truly ϵ-safe, proving the left depth inequality. Conversely, every original-loss terminal state at $\epsilon - 2\eta$ is approximate-terminal at $\epsilon - \eta$; an original row-level policy at the tighter tolerance is a feasible approximate-loss policy. Approximate Bellman optimality proves the right inequality.

For positive capacities n_1, n_2, $\binom{n_1+n_2+2}{2} < \binom{n_1+2}{2}\binom{n_2+2}{2}$, so a genuine merge strictly lowers the product bound. The displayed grid uses at most $\lceil B/(2\delta) \rceil + 1$ values per loss coordinate, and nearest rounding has error at most δ; there are $2M$ coordinates and H outcome strata, giving the stated type bound. □

The corollary does not make the coarsened counts an exact quotient of the original loss problem, give a same-ϵ competitive ratio, or guarantee fewer queries or lower runtime. It certifies a tolerance–state tradeoff; the middle depth is compared with the exact original problem at two different tolerances.

Native-type census and implementation. The structured HEC implementation is checked against explicit labeled worlds on 200 seeded problems with two to seven rows, two to four public types, and two to four candidates. Expected cost and worst depth agree exactly; 92 cases have overlapping decision regions. HEC equals Bayes-optimal expected cost in this census, not in general. An analytic 800-row, two-type family compresses 800 alignments into two count vectors and 801 visited states: exact expected depth is 300.25, worst depth 400, and the realized holder proof has size one. These checks establish compression and proof–search separation, not native CIFAR policy competitiveness. Code, audit records, and regression tests are in the anonymous archive.

The reference implementation is `src/deployrank/public_type_minimax.py`. An explicit-world comparison uses two outcome strata, both signs of context interaction, and $M \in \{3, 4\}$. Across 120 initially unresolved generated releases, every exact minimax depth agrees; the machine-readable record is the author-side path `results/alignment_information_cost/public_type_minimax_audit_v1.json`.

The formulation specializes generic evidence-completion ideas to row identity, a fixed release preserving $P(Y, C)$ and complete $P(Y, Z_{1:M})$, and a regret-defined bad-world family. Large implicit families require separation or problem-specific structure; the paper does not claim polynomial-time exact optimization for every release.

This boundary is concrete on the two prospective acquisition targets. Exact equality of the released outcome and complete candidate-by-context float64 loss vector gives 1,674 public types among 4,000 COIL rows (828 singleton types; maximum capacity 25) and 5,000 singleton types among 5,000 CelebA rows. The direct state bounds $\prod_t (n_t + 1)(n_t + 2)/2$ have respectively 1,305 and 2,386 decimal digits ($\log_{10}$ bounds 1304.33 and 2385.61). Thus the acquisition tables use audited heuristics with the same exact stopping verifier, not the exact Bellman policy. The non-context-using census is recorded in the released public-type audit record. Corollary 9 permits public type merging, but it does not by itself establish scalable planning on these continuous-loss releases: the admissible δ must be small relative to ϵ, and the resulting type census and visited-state count must still be audited.

The prospective zero-acquisition Bank Marketing audit is reported with the decision experiments in Section S2.

G Sampling-based inference and observed alignment

This optional inference extension addresses sampling uncertainty, separately from the exact finite-table certificates in Appendix C. The first results use aligned independent evaluation units; the partial-link results later use a finite-population random query order.

Fix an evaluation object $\mathcal{L}(C, \mathcal{A}, \ell, Q, \mathcal{M})$. For independent evaluation units i, all predictions are frozen before target inference and yield aligned losses $L_{i,0}$, $L_{i,m}$, $L_{i,C}$, and $L_{i,C,m}$. For each

unordered pair (m, n) define

$$G^M_{i,mn} = L_{i,n} - L_{i,m}, \qquad G^V_{i,mn} = L_{i,C,n} - L_{i,C,m}, \qquad G^A_{i,m} = L_{i,C} - L_{i,C,m}. \tag{20}$$

The declared family contains all marginal pair margins, conditional pair margins, and adoption values $V_m(C)$. Repeated rows are reduced to one patient, lesion-component, or molecule unit before inference.

G.1 Declared inference contract and software states

Directional output is conditional on a declared contract fixing the estimand, unit of independence, loss and any clipping bound, candidate stack, coordinate family, and target scope. The command-line interface reports CONTRACT-COMPLETE when these declarations are present and internally consistent, CONTRACT-INCOMPLETE when required metadata are absent, and OUT-OF-SCOPE when the requested claim includes future-domain stability. It never reports "contract validated." A complete declaration permits directional output under the stated assumptions; it does not establish their substantive truth.

The software checks nonempty identifiers, grouping/cluster metadata when declared, duplicate-unit aggregation, nonfinite losses, bounded-loss consistency, prespecified clipping, a frozen candidate stack, a prespecified coordinate family, and unstable standard errors. No observed table can verify substantive independence, absence of hidden clustering, or stability in a future domain. The frozen contract-stress experiment in the *Extended Empirical Evidence* report demonstrates the resulting boundary rather than claiming universal robustness to misspecification.

G.2 Finite-sample bounded-loss certificate

Assume $0 \le L_{i,X} \le B$, independent inference units, non-negative weights w_i summing to one, and K prespecified coordinates. Every paired coordinate lies in $[-B, B]$. Weighted Hoeffding and a union bound give the common radius

$$r_K(\alpha) = B\sqrt{2\left(\sum_i w_i^2\right)\log\frac{2K}{\alpha}}. \tag{21}$$

For equal weights, this reduces to $B\sqrt{2\log(2K/\alpha)/n}$.

Proposition 11 (Simultaneous directional correctness) *On the event that all intervals in the declared family cover their population coordinates, every certified marginal direction, conditional direction, preservation, reversal, adoption, and abstention statement is correct. Consequently, the bounded-loss procedure satisfies*

$$\Pr\{\textit{any false directional certification}\} \le \alpha.$$

Proof. If an interval lower endpoint is positive, coverage implies that the covered coordinate is positive; if its upper endpoint is negative, the coordinate is negative. Preservation and reversal require two such certified directions. The full VALID state requires all relevant pairs to be certified and preserved, whereas INVALID requires a certified reversal. Adoption requires at least one covered $V_m(C)$ lower bound above the prespecified threshold; abstention requires every upper bound below it. A union bound over the K bounded coordinates gives simultaneous coverage at least $1 - \alpha$ using Equation (21). □

Optional variance-adaptive finite-sample mode. For IID equal-weight units, let S_k^2 be the unbiased sample variance of coordinate k. Applying the empirical-Bernstein inequality of Maurer & Pontil (2009, Theorem 4) to each sign and taking a union bound over the K declared coordinates gives simultaneous radius

$$r_k^{\mathrm{EB}}(\alpha) = \sqrt{\frac{2S_k^2\log(4K/\alpha)}{n}} + \frac{14B\log(4K/\alpha)}{3(n-1)}. \tag{22}$$

Proposition 11 then applies unchanged. The software exposes this target-only option as `finite-eb` and fails closed unless the analysis weights are equal and the inference contract attests IID units. It is not used for stratified weights, clustered units, deployment transport, or optional stopping.

A post-hoc exploratory comparison on seven previously specified IID, equal-weight simulation cells finds 2.1–47.4% narrower mean intervals than weighted Hoeffding, with no false direction in 5,000 worlds under either method. Both methods nevertheless remain unresolved in every cell. Thus the variance-adaptive mode improves width in this audit but does not remove the observed finite-sample power boundary.

The result is finite-sample only for independent bounded inference units. For grouped data it applies to independent group means and the declared weighted macro estimand. It does not cover new hospitals, encoder-family sampling, or training randomness absent from the frozen predictions.

G.3 Practical max-$|T|$ certificate

The aligned-data multiplier audit follows the observed-alignment contract introduced at the start of this appendix.

After averaging rows within each declared independent group, let x_{ik} be coordinate k of unit i, with normalized weights w_i and $\hat{\theta}_k = \sum_i w_i x_{ik}$. The implementation uses

$$h = \sum_i w_i^2, \qquad \hat{s}_k^2 = \frac{h}{1-h} \sum_i w_i (x_{ik} - \hat{\theta}_k)^2, \qquad T_b = \max_{k\,:\,\hat{s}_k > 10^{-10}} \frac{|\sum_i \xi_{bi} w_i (x_{ik} - \hat{\theta}_k)|}{\hat{s}_k},$$

where one shared Rademacher vector ξ_b serves every coordinate. The $(1-\alpha)$ empirical quantile of T_b, using the upper order statistic (`method="higher"`), supplies the critical value q for $\hat{\theta}_k \pm q\hat{s}_k$. The simulation uses uniform independent-unit weights; repeated cluster rows are averaged before inference. Thus its grouped estimand equally weights independent-group means, unlike Appendix H's image-weighted whole-component bootstrap. A nonstudentized maximum-deviation sensitivity uses the same draws. Zero or near-zero standard error yields an infinite unresolved interval rather than a zero-radius certificate. Unlike the bounded-loss result, this multiplier procedure is asymptotically justified under regularity conditions for the fixed evaluation stack; it is not claimed to control FWER exactly in every finite sample.

G.4 Partial orders, materiality, and adoption

Certified directions form marginal and conditional directed acyclic graphs. Unresolved edges are retained, the maximal set is reported instead of a forced winner, and a candidate is removed from the ϵ-optimal set only when another candidate is certified better by more than the prespecified ϵ. Directional and material certificates are reported separately.

The no-representation action has $V_\perp(C) = 0$. The adoption axis is ADOPT when some jointly covered lower bound for $V_m(C)$ exceeds the threshold, ABSTAIN when every upper bound is below it, and UNRESOLVED otherwise. Missing resources and infeasible common dimensions are not adoption decisions.

G.5 Conditional-only counterexample

In this two-candidate example, the marginal population values are tied but the conditional values differ. The observed marginal point order therefore does not establish a population direction. Intervals only for the conditional difference can incorrectly label the opposite sign a reversal. The joint procedure leaves the marginal direction unresolved and therefore permits neither a VALID nor INVALID declaration for this example. In a larger menu, another separated pair can still certify global invalidity; one tied pair does not prevent that conclusion. The anonymous software artifact includes an executable regression test.

G.6 GROUPED SUBSAMPLING CALCULATION AND FINITE-SCALE NO-GO AUDIT

Let $W(P)$ denote the common attainable-interval width and let $G(P)$ be the minimum adjacent marginal-value gap in the order frozen before this audit. We target the single joint statistic

$$T(P) = \min\{W(P), G(P)\}.$$

Thus $T(P) > 0$ is exactly the conjunction needed here: the frozen marginal order is strict and the candidate intervals have positive common overlap. This avoids interpreting two separately constructed lower bounds as simultaneous coverage.

Proposition 12 (Grouped-subsampling lower bound) *Suppose source-image groups are i.i.d., group size is uniformly bounded, $b \to \infty$, $b/n \to 0$, and $\sqrt{n}\{T(\widehat{P}_n) - T(P)\}$ converges weakly to a distribution whose $(1-\alpha)$ quantile is continuous. Let $q_{1-\alpha,b}$ be the empirical $(1-\alpha)$ quantile of $\sqrt{b}\{T(\widehat{P}_b) - T(\widehat{P}_n)\}$ over without-replacement complete-group subsamples. Then*

$$L_{n,b} = T(\widehat{P}_n) - q_{1-\alpha,b}/\sqrt{n}$$

has asymptotic one-sided coverage $\liminf_n \Pr\{T(P) \geq L_{n,b}\} \geq 1-\alpha$.

Proof. Under the stated conditions, the subsampling distribution consistently estimates the root-n limiting distribution, so its continuous $(1-\alpha)$ quantile is consistent; inversion yields the displayed lower confidence bound (Politis & Romano, 1994). The assumption is explicit because T includes minima and optimization endpoints: ordinary bootstrap validity is not automatic for nonsmooth directionally differentiable functionals (Fang & Santos, 2016). □

Corollary 10 (Checkable sufficient regularity for the binary log-score audit) *For i.i.d. bounded groups, Proposition 12's root-n and continuous-quantile conditions hold if the population released binary cells are strictly positive, $W(P) > G(P)$, the minimum frozen-order adjacent gap is unique, and that active gap has nonzero cluster influence variance.*

Proof. The row-weighted binary tables are smooth ratios of finite-dimensional group means and hence obey a grouped central limit theorem. On the interior, log-score marginal values are smooth functions of these tables. Uniqueness fixes the active adjacent gap locally. Binary Fréchet bounds and their compact optimization extrema vary continuously with the released tables, so $W(P) > G(P)$ fixes $T(P) = G(P)$ locally. The delta method then gives a nondegenerate normal root-n limit, whose quantiles are continuous. □

The frozen grouped-subsampling calculation uses 1,200 source-image groups, 2,000 draws, and three subsample scales. Although all nominal lower endpoints are positive, the empirical roots remain negatively centered by 1.10–1.52 standard deviations. We therefore do not promote them to population-confidence evidence. A separate 1,000-draw grouped bootstrap finds positive plug-in overlap throughout, while recovery of the complete marginal order is 0.992, 0.802, and 0.596 for log, Brier, and spherical loss. These descriptive checks leave the exact finite-artifact conclusion unchanged.

G.7 RANDOM PARTIAL ALIGNMENT IN A FINITE POPULATION

Fix an N-row target table, a finite set of fixed deployment rules, and a release that exactly fixes $P(Y, C)$ and $P(Y, Z_{1:M})$ while omitting their row alignment. At prespecified looks $b_1 < \cdots < b_K$, reveal a nested simple random sample without replacement. For every candidate pair and relevant projected cell, invert the hypergeometric law and allocate error probability by Bonferroni across cells, pairs, and looks. Intersect the resulting count bounds with the exact released margins to obtain $\mathcal{C}^{\alpha}_{b_k}$.

Proposition 13 (Finite-population partial-alignment selection) *With probability at least $1-\alpha$ over the random row order, the true complete alignment belongs to $\mathcal{C}^{\alpha}_{b_k}$ at every prespecified look. Therefore, if at any such look candidate a satisfies*

$$\max_{Q \in \mathcal{C}^{\alpha}_{b_k}} \left\{ R_a(Q) - \min_j R_j(Q) \right\} \leq \epsilon,$$

then a is ϵ-optimal for the complete finite target table. The same statement holds when stopping at the first successful prespecified look.

Proof. Each inverted hypergeometric interval covers its finite-population cell count at its allocated level. A union bound over the declared cells, pairs, and looks gives one event of probability at least $1 - \alpha$ on which all intervals cover. On that event the true alignment satisfies the release equalities and every count bound, hence belongs to every corresponding confidence polytope. The worst-case inequality applies to it. Simultaneity over looks preserves the claim after stopping at the first successful declared look. □

This fixed-look construction is deliberately elementary. Time-uniform without-replacement confidence sequences permit richer stopping rules (Waudby-Smith & Ramdas, 2021). The following exact error-spending version makes that distinction explicit without changing which row is queried next.

Proposition 14 (Anytime finite-population partial-alignment selection) *Draw one context-blind uniform random permutation of the N target rows. For each member h of a finite family of released strata or projected cells, let $s_h(t)$ be its local sample size after global prefix t. Allocate numbers $\alpha_h > 0$ with $\sum_h \alpha_h \le \alpha$ and, for every $1 \le s < N_h$, invert the corresponding fixed-time hypergeometric law at error*

$$\alpha_{h,s} = \frac{\alpha_h}{s(s+1)}.$$

Use the trivial full-support bound at $s = 0$, the exact count at $s = N_h$, and intersect all current bounds with the released equalities to obtain $\mathcal{C}_t^{\mathrm{any}}$. Then

$$\Pr\{Q^\star \in \mathcal{C}_t^{\mathrm{any}} \text{ for every } 0 \le t \le N\} \ge 1 - \alpha.$$

Consequently, at any stopping time τ adapted to the revealed random prefix, any returned candidate a_τ satisfying

$$\max_{Q \in \mathcal{C}_\tau^{\mathrm{any}}} \{R_{a_\tau}(Q) - \min_j R_j(Q)\} \le \epsilon$$

is non-ϵ-optimal for the complete finite table with probability at most α.

Proof. For a fixed h and local time s, hypergeometric inversion fails with probability at most $\alpha_{h,s}$. Because $\sum_{s \ge 1} 1/[s(s+1)] = 1$, a union bound over every local time and then over h gives simultaneous coverage with failure probability at most $\sum_h \alpha_h \le \alpha$. A uniform global permutation induces a uniform without-replacement order inside every released stratum, so random local sample sizes and interleaving do not change this event. On the event, the true table belongs to every confidence polytope; the displayed robust-regret bound therefore applies at every data-dependent stopping time. □

Corollary 11 (Frozen public priority plus anytime residual sampling) *Before observing any target context, fix a row set A using only the released variables. Reveal every context in A, then draw a context-blind uniform random permutation of the residual rows A^c. Apply Proposition 14 to the residual counts within each released stratum and add the exactly observed counts from A to every lower and upper bound. The resulting full-table confidence sets cover simultaneously over every residual prefix with probability at least $1-\alpha$; hence the same α bound holds for any robust ϵ-selection made at a residual-prefix stopping time.*

Proof. Because A is fixed without target-context access, its revealed counts are constants under the residual randomization. Conditional on this fixed prefix, each stratum of A^c is a fixed finite population and its induced random order is uniform without replacement. Proposition 14 covers all residual counts simultaneously; translating each interval by the known count in A preserves containment and the stopping guarantee. □

The implementation allocates error either over complete aligned cells or over binary-context counts inside complete released prediction strata. Exhaustive audits through population size 12 cover 8,188 binary random-order paths and 90,112 prefixes; a two-stratum audit covers all 630 paths and 4,410 prefixes, hybrid fixed-priority/residual-random audits cover another 630 paths and 4,410 prefixes both for full-stratum count bounds and for the pairwise risk implementation, with no coverage failure at familywise $\alpha = 0.1$. These finite audits check the implementation, while the proposition supplies the general guarantee. The priority set and its size must be frozen before any target context is seen; after that prefix, the next row must still follow the frozen context-blind residual order. Thus nonuniform or response-adaptive row identities require propensity-aware design-based inference and are not validated here. Both propositions concern a fixed finite population, exact linkage, and declared fixed rules, not a superpopulation or refitting.

H OBSERVED-ALIGNMENT SKIN AUDIT: PROTOCOL AND COMPLETE CENSUS

Section S4 gives the result narrative and sensitivity map; this appendix is the single formal protocol and complete target–stack census behind those summaries. It specifies the two observed skin-lesion winner changes reported in Section 5 without requiring readers to infer the protocol from that section. This is protocol self-containment, not a claim that the anonymous ZIP includes the source images, row-level features, or private identifiers. The estimand is the target-domain change in candidate ordering when a frozen source-trained image pipeline is augmented by deployment-time age. The binary outcome is malignant versus benign. HAM10000 and BCN20000 are the two source domains; MSK and PAD-UFES are used only for target evaluation (Tschandl et al., 2018; Combalia et al., 2019; Pacheco et al., 2020). Target labels never select a representation, regularization parameter, calibration map, or context encoding.

Rows, units, and frozen pipeline. The raw one-image-per-lesion manifest contains 7,151 HAM10000 and 3,170 BCN20000 source images. After whole-component duplicate exclusion and the common representation intersection, the four-domain analysis contains 13,523 rows: 10,317 source rows, 1,961 MSK rows, and 1,245 PAD-UFES rows. MSK has 444 positive images and 1,917 patient/duplicate resampling units; PAD-UFES has 832 positive images and 968 units. Source cohorts use lesion-level units because reliable patient identifiers are unavailable there.

Four frozen representations are compared: 124-dimensional handcrafted descriptors, 4,096-dimensional AlexNet features, 768-dimensional DINOv2 features, and 512-dimensional BiomedCLIP features. A PCA map is fit on source features only and retains 16 coordinates for every representation (seed 20260820). Each candidate then uses an L_2-regularized logistic probe. Regularization is selected by leave-one-source-domain-out log loss over $C_{\mathrm{reg}} \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$; the selected inverse-regularization value C_{reg} is reused for that representation's context-plus-candidate probe. Probability calibration is source-only Platt scaling. Context is age dichotomized once at the source-data median. The context-only and context-plus-candidate logistic rules use the same source-only selection discipline. Configuration, representation fitting, and resampling use frozen seed 20260820.

Joint inference and decision rule. For the fitted stacks, the standalone contrast for candidates a, b is $U_a - U_b$ and the context-assisted contrast is $V_a(C) - V_b(C)$, with positive values favoring a. A winner change is reported only when one simultaneous max-$|T|$ family certifies both complete winner statements: the marginal contrast is strictly positive and the conditional contrast is strictly negative. The family contains both targets, both axes, and all six unordered candidate pairs (24 coordinates). Its 20,000 unstratified bootstrap draws sample exactly G_t whole patient/duplicate components with replacement within target t, sharing component multiplicities across every candidate and both axes. Each draw uses the ratio of sampled component loss sums to sampled component sizes, so the estimand remains the image-weighted target-cohort risk, conditional on the frozen source-trained pipelines, while label composition and component size may vary. Independent target streams are paired by draw index. This image-weighted estimand differs from the equally weighted independent-group means used by the multiplier simulation in Appendix G. This is asymptotic simultaneous inference for the two observed target-cohort sampling distributions; it is neither distribution-free coverage nor a claim about an externally representative clinical population.

Because value is risk reduction, $V_a(C) - V_b(C) = R_{C,b} - R_{C,a}$. Thus for a marginal winner a and its displayed conditional winner b, the conditional excess loss of selecting a is the sign reversal $R_{C,a} - R_{C,b} = -(V_a - V_b)$. The reported intervals are exactly these sign-reversed pair-contrast intervals from Table 6. They are not intervals for the selector-regret functional $\max_j(R_{C,a} - R_{C,j})$; a positive pairwise lower endpoint certifies positive selection regret, but its upper endpoint need not upper-bound regret against every candidate.

These are pairwise simultaneous intervals for the two displayed, frozen candidate contrasts; they are not confidence intervals for a data-dependent population-best candidate.

Table 7 gives simultaneous lower bounds for each reported winner against all three competitors. Thus all four winner claims hold within the unchanged 24-coordinate family, while Table 6 highlights the two contrasts demonstrating the winner changes.

Table 6: **Two certified observed-alignment winner changes.** Brackets are simultaneous 95% patient/duplicate-bootstrap intervals from one 24-coordinate family spanning both targets, both axes, and every candidate pair. Contrasts are in nats and are written as marginal winner minus conditional winner.

Target	Pair	Marginal contrast	Conditional contrast
MSK	DINOv2 − BiomedCLIP	$+0.0282$ $[0.0216, 0.0349]$	-0.0111 $[-0.0184, -0.0039]$
PAD-UFES	Handcrafted − BiomedCLIP	$+0.1810$ $[0.1412, 0.2208]$	-0.1308 $[-0.1527, -0.1089]$

Table 7: **Complete winner contrasts from the unchanged simultaneous family.** Entries are simultaneous lower confidence bounds in nats for the reported winner minus each competitor in value. No winner-specific re-criticalization is used.

Target	Axis / winner	Competitor 1	Competitors 2–3
MSK	Marginal / DINOv2	Handcrafted 0.1018	AlexNet 0.0034; BiomedCLIP 0.0216
MSK	Conditional / BiomedCLIP	Handcrafted 0.0176	AlexNet 0.0048; DINOv2 0.0039
PAD-UFES	Marginal / Handcrafted	AlexNet 0.0909	DINOv2 0.0858; BiomedCLIP 0.1412
PAD-UFES	Conditional / BiomedCLIP	Handcrafted 0.1089	AlexNet 0.0306; DINOv2 0.0641

Release-only decision boundary. The preceding comparison asks whether the marginal winner remains best after observed age alignment; it is not the minimax action computed from the unlinked release. A post-hoc sharp top-k audit of the same frozen PCA-16 rules, target outcomes, and released outcome–age counts finds BiomedCLIP weakly dominates every challenger for every compatible row alignment on both MSK and PAD-UFES. Consequently the exact release-only minimax value is $G_0 = 0$ on both targets, as independently re-evaluated with exact binary64 rational arithmetic. The two winner changes therefore establish a failure of marginal selection, but they do not require disclosure to repair the robust release-only decision. The UCI Heart experiment supplies the distinct positive material-repair case.

Complete primary/confirmatory panel. The denominator below was fixed as a heterogeneous audit panel, not sampled from a population of deployment tasks. Sensitivity cells that reuse the same skin targets while changing PCA dimension, seed, projection family, or probe are excluded from the denominator. A zero in the change column means that the point winner did not change; it does not silently upgrade an unresolved joint certificate to certified preservation.

Table 8: **All ten primary/confirmatory observed-alignment audits.** The two positive rows use the joint sign gate above. The other eight show no point winner change, with certification status retained where it is material.

Target	Stack / signal	Marginal → contextual winner	Result
MSK	PCA-16 skin pipeline	DINOv2 → BiomedCLIP	certified change
PAD-UFES	PCA-16 skin pipeline	Handcrafted → BiomedCLIP	certified change
MSK	native skin pipeline	DINOv2 → DINOv2	no point change; unresolved
PAD-UFES	native skin pipeline	BiomedCLIP → BiomedCLIP	certified preservation
ACSIncome 2018	native binary probability	SGD-logistic-onehot → same	no point change
CivilComments test	native binary probability	word–character SGD → same	no point change
iWildCam location target	182-class source-only refit	unchanged	no point change
ImageNet-16H	native prediction stack	unchanged	no point change
CIFAR-10H incumbent	frozen binary superclass	unchanged	no point change
CIFAR-10H native	native ten-class decision	unchanged	no point change

The result establishes occurrence under two target domains of one frozen matched-capacity stack. It does not estimate how often winner changes occur, and the zero native-skin changes show that the observed effect is stack-specific. Broader sensitivity analyses and source-refit diagnostics appear in Sections S4.4 and S4.3, respectively; they are not required to interpret this primary claim. Capacity, seed, and probe variants reuse the same skin labels and are sensitivity cells, not additional occurrence audits; rerunning the primary claim requires the original data and frozen prediction artifacts beyond the submitted ZIP.

I Information contracts and interpretation tables

Table 1 maps the paper's evaluation objects. The two tables here separate the exact inference boundary of each result from the nearest prior work's information and guarantee contract.

Table 9: **Boundary index.** "Exact" refers to the stated finite release; proofs and assumptions appear in the referenced sections.

Established here	Outside this guarantee
For each requested order, one compatible completion realizes it simultaneously for every regular strictly proper loss	A numerical gap uniform over losses, or one completion realizing several incompatible orders
Frozen-table verification and holder frontiers (Appendix C.2); exact joint bounds close where $L = U_{\mathrm{cert}}$	Population guarantees, privacy optimality, or a closed CelebA optimum
Smooth forecasts, stable refits, and sample-based inference under their stated assumptions	Generic calibration, arbitrary representation learning, or deployment prevalence
Exact public-type planning (Appendix F.1) and fixed-K categorical frontiers	Scalable exact planning with proliferating types or polynomial complexity with variable K

Table 10 compares the neighboring work by information available, decision question, and guarantee. In particular, holder-known proof size and online discovery depth are different targets: the native-row

frontier in Theorem 4 addresses the former, while the public-type quotients in Theorems 13 and 12 address structured versions of the latter.

Table 10: **Nearest-work comparison by information, quantifier, and guarantee.**

Closest object	Information, question, and distinction
Matching/PID; data-fusion DML (Cross & Manski, 2002; Ahfock et al., 2016; Bertschinger et al., 2014; Jiang & Janson, 2025)	*Information:* lower-order marginals or independent data sources. *Question:* bound scalar functionals or efficiently estimate moment-based outer endpoints. *Here:* fixed candidate-wise marginals yield a sharp Cartesian M-candidate value image; a separate complete-joint release admits local all-order shattering.
Structured completion (Koo et al., 2026)	*Information:* marginals plus a compatibility regularizer. *Question:* select one assumption-indexed coupling. *Here:* certify decisions uniformly over every release-compatible completion instead of choosing one.
Deployment-complete benchmarking Mansouri & Arai (2026)	*Information:* an abstract evidence map and its fibers. *Question:* decide whether an action is constant; acquire probes to refine evidence. *Here:* fixed $P(Y, C)$ and complete $P(Y, Z_{1:M})$ still permit local all-$M!$ shattering; a sharp adequacy test and quota-constrained row-link frontier certify fixed-rule choices.
HEC / decision-region determination Javdani et al. (2014)	*Information:* finite hypotheses, prior, tests, and overlapping decisions. *Question:* minimize expected test cost with a Bayesian competitive greedy policy. *Here:* an exact public-type quotient preserves HEC queries under a rational count prior with uniform within-type labelings; its bound is inherited, while prior-free results address different costs.
Constrained sufficient reasons (Barceló et al., 2020; Gorji & Rubin, 2021)	*Information:* one input with coordinate-level completion restrictions. *Question:* fix a smallest coordinate set that preserves its verdict over allowed completions. *Here:* the shared outcome–context quotas couple row labels, and one selected candidate must satisfy all challenger inequalities at once. The pairwise frontier and joint-selection bounds price authenticated row links; online discovery is a separate query-policy problem.

S1 Extended Empirical Evidence: role and scope

The experiments evaluate three distinct consequences of missing alignment. Compatible-ranking studies test whether one released prediction table admits different contextual orders. Decision studies test whether the unlinked release already certifies an ϵ-optimal candidate and, if necessary, how many disclosed links complete that certificate. Observed-alignment studies assess marginal and contextual comparisons under a declared sampling and fitting protocol. These evidence types answer identification, decision, and inference questions, respectively. Real-data executions are author-controlled audits of historical datasets; separate implementations establish replayability, not independent custody or clinical validation. The fixed study panel does not estimate deployment prevalence. Table S3 classifies decision outcomes, and Table S22 specifies which inputs permit independent replay. Unless stated otherwise, empirical upper bounds are verified certificate sizes U_{cert}, not the union-restricted bound U^{union}.

For observed-alignment inference, a trusted evaluator receives aligned entity-level losses and uses one simultaneous family for three axes: **leaderboard validity** (certified, consistent marginal and context-assisted pair directions), **adoption** (certified value beyond no representation), and **transport** (persistence of source-conditional directions on target). Axes remain separate; a non-significant pair is UNRESOLVED.

S2 Decision and disclosure evidence: How much alignment resolves selection?

S2.1 Compatible-ranking locality

Figure S1 measures which rankings remain compatible as a TV restriction around the observed completion is relaxed.

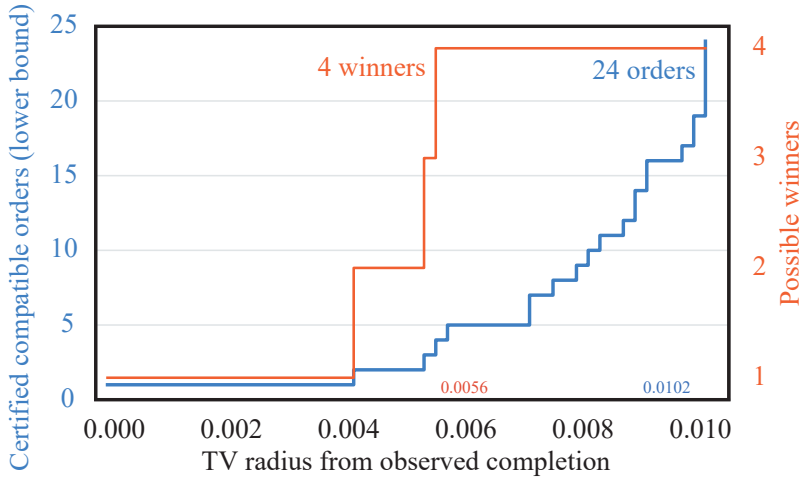

Figure S1: **Restriction sensitivity on the frozen CIFAR-10H witness.** Step curves are certified lower bounds inside TV balls around the observed completion, rather than interpolated identified-set sizes. Every candidate is a possible winner by radius 0.0056; all 24 log-score orders are witnessed by 0.0102.

Table S1: **Fixed-artifact witnesses.** CXR: all-loss; others: score-wise; native: post-hoc fixed rules.

Evidence	M	Signal	Released prediction law	Primary result / status
ImageNet-16H (I16-JOINT)	5	binary	complete $P(Y, Z_{1:5})$	120/120 log; frozen reuse
CIFAR-10H (C10-JOINT)	4	binary	complete $P(Y, Z_{1:4})$	24/24 log; post-result
CXR	5	binary	complete $P(Y, Z_{1:5})$	120/120 Blackwell
Adult	5	binary	complete $P(Y, Z_{1:5})$	120/120 log; result-blind
ImageNet native (I16-NATIVE)	4	16-class	complete $P(Y, \mathbf{P}_{1:4})$	24/24 per score; post-hoc

S2.2 Exploratory CIFAR acquisition

An execution-frozen CIFAR-10H acquisition audit starts from a previously opened but fully aligned finite table. A deterministic row-index split uses 5,000 rows to fit four Jeffreys-smoothed fixed log-loss rules and 5,000 target rows for evaluation. The target release retains exact $P(Y, C)$ and complete $P(Y, Z_{1:4})$ but hides their row linkage. Each repetition draws one nested uniform target-row permutation; revealed rows become hard lower bounds in the release fiber. The method returns a candidate only when its sharp worst-case regret over every remaining completion is at most the prespecified $\epsilon = 0.005$ nats. This finite-table guarantee is valid for any correctly revealed subset and is distinct from population sampling coverage. The comparison treats the released outcome–context counts as pre-existing. If producing them requires labeling every target context, our row count prices only later linkage revelation or verification, not that upstream labeling cost.

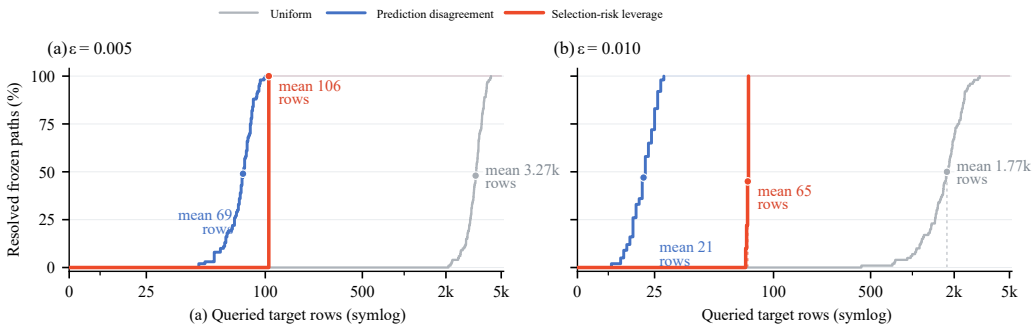

Figure S2: **Exact acquisition depends on which public rows are checked.** All three policies use the same sharp release fiber and 100 paired paths. Priority scores never use unrevealed target context. The comparison is exploratory.

Two frozen statistical tracks use the same 100 nested random paths: simultaneous hypergeometric bounds either stand alone or intersect the release fiber at nine prespecified looks. Proposition 14 and Corollary 11 give the anytime and public-prefix versions. Implementation audits cover 8,188 binary paths and 1,890 stratified/hybrid paths without failure at $\alpha = 0.1$. These are sampling-design guarantees; the targeted priorities below retain deterministic hard-fiber stopping.

Table S2: **Exploratory sampling-aware comparison at scheduled looks.** Entries are mean first successful scheduled-look row counts over 100 frozen nested paths, not the later per-row first-stop counts; 5,000 means every path first resolves only at full alignment. All paths have simultaneous interval coverage and no false ϵ-certificate.

Method	$\epsilon = 0.005$	$\epsilon = 0.010$
Hard release fiber	5,000	2,960
Aligned sample only	5,000	5,000
Release + sampling bounds	5,000	2,950

Anytime bounds were valid but changed none of the 600 hard-fiber stopping decisions. At exploratory $\epsilon = .010$, the hard-fiber and hybrid means differ by only ten rows, while sample-only inference waits for full alignment. The targeted policies below use deterministic fiber certificates; their query identities, not the sampling overlay, drive the larger savings. The author-side audit retains the path-level table and independent 13,200-decision reconstruction; the submitted ZIP does not provide full path replay.

Release-only inference keeps all four candidates possible winners; the original uniform-query grid remains unresolved through 2,000 rows and resolves at 5,000. A disclosed 30-path prototype fixed two release-visible priorities and a finer grid before the 100-path comparison. A later per-row bisection audit tightens the grid upper stops. At $\epsilon = 0.005$, mean exact first stops are 69.10 rows for candidate-output disagreement, 106 for selection-risk leverage, and 3,270.02 for uniform order; at 0.010 they are 21.42, 65.23, and 1,771.99. Both priorities beat uniform on every paired path with no false certificate. The 122 rows with any candidate-output disagreement are public before acquisition; risk leverage likewise uses only the conditional-independence incumbent and source-fitted losses. Hidden context enters only when a selected row is queried. Simple disagreement is best here, supporting release-visible targeting but not superiority of the more elaborate score; neither priority has a proved approximation ratio. On one opened path at $\epsilon = 0.010$, post-hoc HEC selects candidate 3 in 10 queries versus disagreement's 23 (true regret 0.000430). Across 100 six-row problems it matches exact minimax depth in 82; no approximation guarantee transfers to the implicit fiber.

Appendix F.1 reports the structured-release HEC implementation checks separately; they do not establish competitiveness of these CIFAR query priorities.

Rows matching in outcome, true context, and complete loss table form 35 exact types. At $\epsilon = 0.005$, exhaustive search covers 531,160,616 states across the closed eight- and nine-row layers. Candidate 3's best nine-row risk is $0.005102408 > 0.005$; Proposition 3 excludes every ninth row for candidates 1 and 2. Here $B_a^{(8)}$ is the smallest worst-case regret among all eight-row queried sets for candidate a, and δ_a is a sequence-uniform upper bound on the improvement produced by any feasible ninth query. Thus $B_a^{(8)} - \delta_a$ is a lower bound over *every* nine-row set, not the value of one selected extension. The rounded $(B_a^{(8)}, \delta_a, B_a^{(8)} - \delta_a)$ triples are $(0.019663, 0.002382, 0.017281)$ and $(0.012461, 0.002325, 0.010136)$. The ten-row witness has risk $0.004984315 < 0.005$. Full-precision ledgers cover every eight-row set and possible ninth row; all feasible sets are rechecked on the original 5,000 rows. Exhaustive search through budget four plus a five-row witness similarly closes $\epsilon = 0.010$. Thus

$$\kappa^{\text{select}}_{0.005} = 10, \qquad \kappa^{\text{select}}_{0.010} = 5.$$

Because the types use opened truth, these are post-hoc costs, not online-policy results or approximation guarantees for the reported query priorities. When every bad world is explicitly enumerated, greedy row cover has the standard $H_{|\mathcal{B}|} \le 1 + \log|\mathcal{B}|$ guarantee. That result does not make the public-disagreement order near-optimal here: this 5,000-row fiber is implicit, and the reported policy is not greedy cover over an enumerated bad-world family.

Conditional-independence completion selects DenseNet, but complete alignment selects PreResNet and gives DenseNet regret $0.000430 < 0.005$. Continuous-law TV radii are $0.001107/0.001506$; exact lattice optima move 6/8 rows (TV $0.0012/0.0016$). Nearby bad worlds and acquisition costs are therefore different objects. All priorities were developed on opened data, not as independent validation or frequency evidence.

S2.3 Cross-setting decision regimes

Table S3 separates zero-link sufficiency, guarantee repair, and above-tolerance choice repair, while reporting holder proof bounds separately from online query counts.

Reading the task records. Each study states its information contract, protocol status, observed decision, and certificate cost. Execution-only amendments and entry points are collected in Section S5.

Bank Marketing and Online Shoppers certify their frozen winners at zero queries for every threshold (Moro et al., 2014; Sakar et al., 2018). For Bank Marketing, source models, release-visible policies, budgets, thresholds, and an independent auditor were frozen before opening the unused 9,188-row chronological target. Its exact release fiber already gives one candidate zero worst-compatible regret. Every policy therefore stops at zero for both declared positive thresholds, and complete alignment selects the same candidate. Independent code reconstructs 4,500 budget bounds and 9,000 threshold decisions without a false certificate. This is a fail-closed boundary result, not evidence of a targeted-query advantage.

FMoW uses 22,108 test images, revision-pinned 62-way forecasts from four models, four one-versus-rest regions, and year ≥ 2017. Its fixed source-only rule multiplies each forecast by $P(C \mid Y)/P(C)$ (Laplace 0.5), without target fitting or calibration; all ten context–threshold cells certify CLIP-L/14 and complete alignment agrees (Koh et al., 2021). Table S4 applies one diagnostic to these releases and the nonzero-query examples. At zero tolerance, FMoW remains box-safe: its five reused context settings in Table S4 have fiber leads 0.08798–0.14754 nats. Bank and Online require released counts, whereas CIFAR and Waterbirds require alignment. The FMoW year row is extrapolative: source has 88,346 context-zero rows and 0 positives, versus 6,149 target positives.

S2.4 Waterbirds: proof size and discovery cost

The frozen static priorities have exact, shared definitions. For candidate positive-class probabilities p_{im}, Bernoulli–JS assigns row i

$$J_i = h(\bar{p}_i) - M^{-1}\sum_{m=1}^{M} h(p_{im}), \qquad h(p) = -p\log p - (1-p)\log(1-p),$$

where $\bar{p}_i = M^{-1}\sum_m p_{im}$ and probabilities are clipped to $[10^{-12}, 1-10^{-12}]$. Selection-risk leverage fixes the release-only incumbent a_0, sets $\Delta_{mi} = \ell_{mi1} - \ell_{mi0}$, and scores $L_i = \max_{j \ne a_0} |\Delta_{a_0 i} - \Delta_{ji}|$.

Table S3: **Cross-setting decision and evidence ledger at $\epsilon = .005$.** Panel A separates a wrong action from a missing guarantee; Panel B separates a truth-known holder proof from online discovery. Opened CIFAR is exploratory. COIL/CelebA minimax entries are post-hoc exact replays, although their acquisition protocols preceded target opening. Design codes are E (exploratory), FP (target-frozen with a post-hoc planner), and P (protocol-prospective); H marks post-hoc minimax accounting. P* marks a separately timestamped protocol with a disclosed source-informed amendment; only its amended target stage was sealed before opening. Repeated paths on one release are not independent target tasks.

A. Release-only decision consequence

Setting	G_0	MM action → aligned action	Regret	Design	Effect
Opened CIFAR-10H	.01588	DenseNet → PreResNet (post-hoc)	.00043	E	point winner
Waterbirds val	.24276	$2 \to 2$	0	FP	guarantee
Waterbirds test	.24028	$2 \to 2$	0	FP	guarantee
COIL-2000	.00676	$0 \to 0$	0	P+H	guarantee
CelebA	.04098	$2 \to 2$	0	P+H	guarantee
Heart H–S	.05684	RF → HGB	.00790	P+H	choice + guarantee
Heart H–VA	.05155	HGB → HGB	0	P+H	guarantee
Heart S–VA	.07207	HGB → HGB	0	P+H	guarantee
I-CARE A–B	.01669	ExtraTrees → ExtraTrees	0	P*	guarantee
I-CARE E–F	.03705	HGB → HGB	0	P*	guarantee
UCI Diabetes	.00290	ExtraTrees → ExtraTrees	0	P	zero-link
PTB-XL MI/age	0	HGB → HGB	0	P	zero-link
UCI Sepsis N–K	0	HGB → HGB	0	P	zero-link

B. Truth-known proof versus online discovery

Setting	Holder L/U_{cert}	Priority	Mean links: priority / uniform
Opened CIFAR-10H	10/10	prediction disagreement	69.10 / 3,270.02
Waterbirds val	123/123	expected-gap	184 / 606.96
Waterbirds test	619/619	expected-gap	995 / 2,825.81
COIL-2000	159/159	selection-risk leverage (secondary)	335.81 / 911.04
CelebA	263/302	Bernoulli-JS disagreement (secondary)	1,019 / 3,842.56
Heart H–S	86/86	Bernoulli-JS disagreement	278 / 361.72
Heart H–VA	71/71	Bernoulli-JS disagreement	212 / 327.54
Heart S–VA	69/69	Bernoulli-JS disagreement	151 / 239.58
I-CARE A–B	18/25	expected-gap (secondary)	57.48 / 99.38
I-CARE E–F	4/4	expected-gap (secondary)	11 / 20.47
UCI Diabetes	0/0	Bernoulli-JS disagreement	0 / 0
PTB-XL MI/age	0/0	Bernoulli-JS disagreement	0 / 0
UCI Sepsis N–K	0/0	Bernoulli-JS disagreement	0 / 0

Table S4: **Observable geometry separates zero- and nonzero-query releases.** Signed margins are positive only when one candidate is strictly optimal throughout the uncertainty set; G is minimum worst-case regret. Regime labels use zero tolerance; at tolerance ϵ, alignment is required exactly when $G_{\text{fiber}} > \epsilon$. The box drops outcome–context counts. Values are rounded to five decimals.

Setting	Fiber margin	Box margin	G_{fiber}	Mean $\lvert\Delta\rvert$	Regime
Bank	0.04472	-0.13989	0.00000	0.30068	count-rescued
Online	0.00014	-0.00781	0.00000	0.06357	count-rescued
FMoW Africa	0.14754	0.07774	0.00000	0.29917	box-safe
FMoW Americas	0.08798	0.05727	0.00000	0.28882	box-safe
FMoW Asia	0.10908	0.05944	0.00000	0.31086	box-safe
FMoW Europe	0.11026	0.09396	0.00000	0.23051	box-safe
FMoW year ≥ 2017	0.12722	0.11339	0.00000	0.28137	box-safe
CIFAR-10H	-0.01588	-0.01594	0.01588	0.33824	alignment-required
Waterbirds val	-0.24276	-0.27027	0.24276	0.84656	alignment-required
Waterbirds test	-0.24028	-0.26676	0.24028	0.86544	alignment-required

Table S5: **Waterbirds acquisition protocol.** Both targets and all outcomes are retained; all policies share the exact full-fiber verifier.

Component	Frozen specification
Data	train source; validation (1,199) and test (5,794); hidden binary background
Rules	four revision-pinned native probabilities; source-only fusion, Laplace 0.5
Threshold / stop	$\epsilon \in \{.005, .010, .020\}$; exact worst remaining regret
Policies	post-hoc expected-gap; disagreement; risk; uniform (100); sampled HEC (5)

Both query larger scores first and break ties with the frozen, context-independent key; uniform uses the corresponding frozen random permutations.

For current fiber $\mathcal{F}_t$, write $G(\mathcal{F}) = \min_a \max_{Q\in\mathcal{F}}\{R_a(Q) - \min_j R_j(Q)\}$. Here $r^{(t)}_{y,c}$ is the residual released count assignable to context c within outcome stratum y, $u^{(t)}_y$ is the number of unrevealed rows in that stratum, and $\mathcal{F}^{i,c}_t$ is the child fiber after assigning row i the answer c. Expected-gap scores every unqueried row by $s_t(i) = \sum_{c:r^{(t)}_{y_i,c}>0}[r^{(t)}_{y_i,c}/u^{(t)}_{y_i}][G(\mathcal{F}^{i,c}_t) - \epsilon]_+$. Pseudocode: compute feasible children; minimize s_t; break ties by expected child G, worst child G, then public row ID; reveal, recompute, and stop only when exact $G(\mathcal{F}_t) \leq \epsilon$. Count fractions are an exchangeable planning law, not a context model. Each exhaustive step costs $O(M^2N\log N)$ time and $O(M^2N)$ memory; no approximation guarantee is claimed.

Table S6: **Waterbirds: discovery cost versus the exact holder certificate.** Uniform/HEC are 100-/5-path means; HEC has sample SD 0 in all six cells. Test baselines extend the 1,000-row cap to exact stop. The last column is the minimum sufficient disclosure, not an online policy.

Split	ϵ	Exp.-gap	Disagr.	Risk	Uniform	HEC	Exact holder
validation	.005	184	395	478	606.96	664 ± 0	**123**
validation	.010	173	378	458	590.36	654 ± 0	**120**
validation	.020	163	353	442	562.58	630 ± 0	**113**
test	.005	995	2,045	2,556	2,825.81	$2{,}796 \pm 0$	**619**
test	.010	965	1,992	2,505	2,762.53	$2{,}709 \pm 0$	**600**
test	.020	902	1,911	2,401	2,633.11	$2{,}603 \pm 0$	**564**

All six cells use the strict event $\text{regret} > \epsilon$ and stop at exact remaining-fiber regret at most ϵ (tolerance 10^{-12}). JSON indices are zero-based; index 2 is CLIP-L/14, selected under complete alignment. For every cell, its hardest pairwise disclosure minimum is also a valid certificate against all other challengers, proving the displayed global selection optimum. Expected-gap requires 1.44–1.61 times

this truth-known minimum. That ratio measures discovery overhead on these realized alignments, not a worst-case approximation guarantee for the online policy.

On the test split, expected-gap uses 47.2–48.7% as many queries as disagreement and 34.3–35.2% as many as uniform, while every policy uses the same shared exact stopping certificate. A separate author-run implementation reproduces the planning decisions and first stops. Expected-gap requires more computation: optimized median runtimes are 13.9–15.3 seconds, compared with 2.28–2.45 seconds for disagreement and 2.88–3.07 seconds for risk priority. Under a homogeneous serial reveal-cost model, it becomes cheaper than disagreement at approximately 11.6–12.2 ms per reveal. These are same-host computational and break-even diagnostics, not measurements of human or governance cost; the hardware, software, and threshold-specific values are recorded in the reproducibility ledger below. This query–compute tradeoff implies that above the displayed break-even costs, planning is not computationally dominant in the combined plan-plus-reveal calculation. The separate reimplementation is an author-run software check, not external human review.

S2.5 Protocol-prospective nonzero-query targets

COIL-2000 was not used for method development. Before parsing its official 4,000-row evaluation table, we froze the source split, four demographic-only native-probability candidates, car-policy context, likelihood-ratio fusion, $\epsilon = 0.005$, 100 tie-breaking seeds, policies, and a no-rescue rule (P. van der Putten and M. van Someren, *The Insurance Company Benchmark (COIL 2000)*, UCI KDD Archive, 2000). The target release is initially deficient: $\mathfrak{G} = 0.00676254$. Complete alignment selects logistic regression at risk 0.213751; every reported policy stop has true regret zero.

Table S7: **COIL-2000: prospective exact acquisition.** Dispersion is across 100 frozen tie-breaking paths on one target, not 100 independent targets; runtimes are per path. The 25% expected-gap advantage over both informed baselines was the primary efficacy gate.

Policy	Links: mean±SD	IQR	Time (s)	Saving
Expected-gap	336.62±0.49	336–337	2.309	63.1%
Bernoulli-JS	340.40±0.49	340–341	0.192	62.6%
Selection-risk	**335.81±0.39**	336	**0.190**	**63.1%**
Uniform	911.04±46.80	880.75–944.50	0.520	—

The expected-gap dominance gate did not pass. The exact frontier closes the truth-known minimum at 159 rows: pairwise costs are 145, 159, and 31, and the bottleneck certificate protects against all challengers. The best frozen policy uses 335.81 rows on average (2.11× the holder proof size). A non-importing audit reconstructs all 400 policy paths (four policies by 100 tie-breaking paths), including their individual answers and first stops, with zero false certificates under the prospective acquisition protocol.

Real-scale categorical disclosure stress. Post-hoc reuse of the four probability columns with $C = \min(\texttt{APERSAUT}, 2)$ gives $71 \leq \kappa^{\text{select}}_{.005} \leq 238$. A public-only exact verifier checks the 238-link (5.95%) certificate. It exhausts the 152-row third category, so this is realistic-size categorical sufficiency, not generic optimizer scaling.

CelebA provides a second target opened only after its separate protocol was frozen and after the COIL result was known (Z. Liu et al., *Deep Learning Face Attributes in the Wild*, ICCV, 2015). The target is a hash-selected 5,000-image subset of the official test partition; Y is the dataset's `Blond_Hair` annotation and C is its binary `Male` annotation, which we do not interpret as gender identity. Four ImageNet-pretrained encoders feed full-dimensional source-only logistic heads, with Platt calibration on the official validation split. No target row enters training or calibration. The release is initially deficient ($\mathfrak{G} = 0.0409755$ at $\epsilon = 0.005$), and complete alignment selects EfficientNet-B0 at risk 0.154760.

Table S8: **CelebA: frozen native-image exact acquisition.** Dispersion is across 100 frozen tie-breaking paths on one target, not 100 independent targets; the primary expected-gap superiority gate was fixed before target opening.

Policy	Links: mean±SD	IQR	Time (s)	Saving
Bernoulli-JS	**1,019.00±0.00**	1,019	**1.192**	**73.5%**
Selection-risk	1,121.00±0.00	1,121	1.297	70.8%
Expected-gap	1,191.06±71.56	1,131.25–1,257.50	16.091	69.0%
Uniform	3,842.56±104.15	3,762.25–3,908.00	5.009	—

All 400 paths certify and an implementation-independent sorter reconstructs every individual answer, preceding unresolved state, first stop, and query hash, with zero false certificates. The prespecified expected-gap dominance gate did not pass; the robust online finding is instead a 62.6–73.5% reduction for release-informed priorities relative to uniform querying on two fresh, initially deficient targets. Holder-side frontiers instead give $263 \le \kappa^{\mathrm{select}}_{.005} \le 302$ and $174 \le \kappa^{\mathrm{select}}_{.010} \le 195$. Lower endpoints are exact pairwise floors over every truth-feasible candidate; upper endpoints are unrestricted joint-MILP certificates rechecked in exact dyadic arithmetic. The 5,000-binary runs took 2403.4/276.7 seconds and reported zero numerical gaps, but solver status is not exact proof; only the displayed rigorous brackets are claimed.

S2.6 HEART: MATERIAL CHOICE REPAIR

Before downloading the official UCI Heart Disease archive (DOI 10.24432/C52P4X), we froze Cleveland-only training, four candidates, three pairs of unseen hospitals, $\epsilon = .005$, one primary Bernoulli-JS priority, 100 tie-breaking paths, and a no-rescue rule. Hospital identity is the hidden context inside each target pool; target outcome–hospital counts supply the fixed likelihood-ratio context update. The overlapping pairs are three settings of one cross-hospital family, not three independent target studies. The official UCI archive supplies Cleveland ($n = 303$) as source and Hungary, Switzerland, and VA Long Beach as targets; $Y = \mathbf{1}\{\mathrm{num} > 0\}$. Candidates use the 13 standard fields except sex, source-median imputation, and probability clipping at 10^{-4}: L2 logistic ($C = 1$), 500-tree RF/extra trees (leaf 5, sqrt features), and histogram boosting (300 iterations, rate .05, 15 leaves, L2 penalty 1). Five-fold stratified Cleveland OOF predictions (seed 450020) are Platt-calibrated before all-source refits (candidate seeds 450021–450024). The fixed context update uses Jeffreys 0.5 smoothing. Evaluation uses exact finite-table log loss, 100 orders from seed 20260920, and common free-quota closure in the post-hoc accounting. Hashes and execution entry points are in Section S5.

Table S9: **UCI Heart under the original frozen accounting.** JS/uniform entries are mean±sample SD links over 100 tie-breaking paths; all JS paths are identical because its scores have no boundary ties. "Ref." is the release-only conditional-independence choice. Candidate indices $0/1/2/3$ denote logistic regression, random forest, extra trees, and histogram gradient boosting.

Pair	N	Initial $\mathfrak{G}$	Ref.→aligned	Ref. regret
H–S	417	.056843	$1 \to 3$	.007900
H–VA	494	.051547	$3 \to 3$	0
S–VA	323	.072074	$1 \to 3$	.016828

Pair	JS links	Uniform links	Reduction
H–S	278 ± 0	365.08 ± 13.10	23.9%
H–VA	212 ± 0	327.54 ± 14.67	35.3%
S–VA	151 ± 0	240.01 ± 10.78	37.1%

The decision-impact gate above used the frozen conditional-independence reference. A post-hoc audit also compares the paper's stronger release-only minimax-regret action, closes the truth-known holder proof, and applies the same free quota closure to every path:

Table S10: **Heart under post-hoc common free-quota accounting.** The same frozen JS stopping paths are retained; common quota-forced answers change only the paid-link accounting shared by all policies. G_0 is public worst-compatible regret; realized regrets use hidden alignment only for evaluation. L/U are holder proof bounds and all three close.

Setting	CI: action/regret	MM: action/regret	G_0
H–S (417)	RF→HGB/.00790	RF→HGB/.00790	.05684
H–VA (494)	HGB→HGB/0	HGB→HGB/0	.05155
S–VA (323)	RF→HGB/.01683	HGB→HGB/0	.07207

Setting	Proof L/U_{cert}	JS / uniform	Saving
H–S (417)	86/86	278/361.72	23.1%
H–VA (494)	71/71	212/327.54	35.3%
S–VA (323)	69/69	151/239.58	37.0%

The primary decision-impact gate, defined against the frozen CI reference, passes on two settings and the primary efficiency gate passes on all three. The secondary expected-gap policy uses 174/139/128 links, but the transfer claim is based on the predeclared JS policy, not retrospective policy selection. A direct sorter independently verifies all 1,200 policy paths (four policies by 100 paths across three settings), their individual answers, preceding unresolved states, first stops, and query hashes with zero false certificates. The pre-query cap amendment changed no scientific setting; execution details are recorded in Section S5.

Process-separated query replay (post hoc). On Hungary–Switzerland, an evaluator process loaded only the public release; a separate answer service loaded the hidden row–hospital map and returned one link per request. The evaluator checked the response chain and independently recomputed the exact binary64-rational certificate: the 277-link prefix remains unresolved (minimum worst regret .005278), while query 278 certifies HGB ($.004391 < .005$). Query and answer hashes match the frozen JS path. The aggregate-only record is produced by `audit_heart_holder_evaluator_replay_v1.py`. Both processes ran under author control on one machine; the chain is unsigned and does not authenticate the source records or establish an external holder.

Post-hoc information-contract sensitivity. The base contract releases outcome–hospital counts but no row–hospital links. We additionally audit two nested, candidate-independent contracts: T1 releases row-visible recorded sex and $Y \times S \times C$ counts; T2 also releases the fixed decade $A_{10} = \lfloor \mathrm{age}/10 \rfloor$ and $Y \times S \times A_{10} \times C$ counts. No model, loss, target, ϵ, or bin boundary changes. Sparse strata are retained, and quota-forced identities are inferred before charging a query. T0 is therefore rerun with the same free quota closure; its uniform means can differ slightly from Table S9, while the fixed JS stops do not.

All nine pair–tier releases remain unresolved at $\epsilon = .005$; the two incorrect CI-reference choices and their $.007900/.016828$ regrets persist.

T2 still contains $2^{197.1}$–$2^{261.9}$ compatible assignments after 13.4–20.9% of rows are quota-forced in two pairs and 1.9% in the third. Thus coarse demographic aggregates narrow but do not remove the need for replayable evidence in these instances. Conversely, a custodian that publishes four complete-alignment risks rounded to 10^{-6} certifies all three choices with zero row links. That comparator requires trusting the hidden-data calculation: hashes identify inputs but do not make the four values replayable from the unlinked release. The analysis was designed after the Heart results were known and is contract sensitivity, not new prospective evidence or a privacy claim. An independent direct-sorter implementation replays all 1,800 tier–policy paths.

Post-hoc tolerance and rule-menu robustness. The saved five-tolerance audit uses the original losses, quotas, and policies; an independent implementation checked its 3,000 paths. H–S remains a material minimax choice error at $.0025/.005/.0075$, but at $.010/.020$ its .00790-nat error lies within tolerance although the release remains uncertified. The first three holder minima are exact; the higher-tolerance results are bounds, not exact minima (Table S12).

Table S11: **Heart information-contract sensitivity (post hoc).** T0 is the original Y contract; T1 adds recorded sex; T2 adds fixed age decades. $\mathfrak{G}$ is the initial minimum worst-case regret, "Forced" is the fraction of hospital identities implied by public quotas before queries, and query columns are means over the same 100 fixed orders. The settings share hospitals and are not independent target studies.

Pair	Tier	$\mathfrak{G}$	Forced	$\log_2 \lvert\mathcal{F}\rvert$	JS/uniform	Saving
H–S	T0	0.05684	0.0%	261.9	278/361.72	23.1%
H–S	T1	0.05390	16.5%	254.0	229/300.57	23.8%
H–S	T2	0.04663	20.9%	197.1	174/242.52	28.3%
H–VA	T0	0.05155	0.0%	420.2	212/327.54	35.3%
H–VA	T1	0.04783	0.0%	392.9	205/313.98	34.7%
H–VA	T2	0.02434	13.4%	261.9	110/181.42	39.4%
S–VA	T0	0.07207	0.0%	287.5	151/239.58	37.0%
S–VA	T1	0.07134	0.9%	280.4	140/235.77	40.6%
S–VA	T2	0.06315	1.9%	248.4	131/212.27	38.3%

We also evaluate three fixed menus without refitting: four original fused rules; those four plus their unfused forecasts; and those eight plus a context-only rule $q_c = (n_{1c} + 1/2)/(n_{0c} + n_{1c} + 1)$ computed from public counts. These are separate scans: 15 original-menu tolerance cells and nine 4/8/9-rule menu cells at $\epsilon = .005$, not a 45-cell crossed design. H–S keeps fused RF as the minimax action and fused HGB as the aligned winner, with .00790 regret across all three menus. H–VA keeps fused HGB and zero aligned regret. S–VA becomes zero-link certified after adding context-only. Thus H–S repair survives these natural comparators, while another release becomes sufficient: the menu is part of the evaluation contract, not an innocuous reporting choice. Expanded-menu proof sizes and query costs are not inferred from the original 86-link proof or frozen trajectories. For H–S, G_0 rises from .05684 to .06192 to .10285: expanding the menu adds both available actions and adversarial comparators inside regret, so its minimax value need not decrease.

Table S12: **Heart robustness on retained inputs (post hoc).** Top: H–S tolerance sensitivity under the original four rules. Bottom: all three 4/8/9-rule menus at $\epsilon = .005$; G_0 is minimum worst-compatible regret and aligned regret is that of the minimax action. Brackets are bounds; single holder counts are exact. All five tolerance cells still require linkage: the last two are guarantee-only repair, not zero-link certification. RF/HGB denote fused rules.

H–S tolerance	.0025	.005	.0075	.010/.020
Repair type	Choice	Choice	Choice	Guarantee only
Holder bounds	91	86	82	[65,78]/[40,58]
Menu / target	Rules	G_0	Aligned regret	Action
H–S	4	0.05684	0.00790	RF-fused
H–S	8	0.06192	0.00790	RF-fused
H–S	9	0.10285	0.00790	RF-fused
H–VA	4	0.05155	0.00000	HGB-fused
H–VA	8	0.06417	0.00000	HGB-fused
H–VA	9	0.06417	0.00000	HGB-fused
S–VA	4	0.07207	0.00000	HGB-fused
S–VA	8	0.07207	0.00000	HGB-fused
S–VA	9	0.00000	0.00000	Context

An existing complete source-rotation census evaluates all four possible source hospitals and all three pairs among each source's remaining hospitals: 3 of 12 settings have above-.005 minimax choice error. This exploratory census reuses the same hospitals, is not an independent task family, and does not establish three completed acquisition repairs. All 15 tolerance cells, 12 rotations, and nine menu cells are supplied in `examples/heart_replay/sensitivity_summary.json`. H–S menu values can be recomputed from the supplied public losses/probabilities and, separately, holder alignment using `menu_sensitivity.py`; the other menu and rotation summaries do not supply

full input replay. These checks assess sensitivity after results were known, not additional prospective evidence.

S2.7 I-CARE: FROZEN CROSS-TASK TRANSFER

To test whether acquisition behavior transfers beyond the development tasks, we timestamped an I-CARE v2.1 protocol before opening target metadata (Amorim et al., 2023). A deterministic patient-hash split assigns 340 patients to source fitting and 267 to held-out targets. The binary task is $Y = \mathbf{1}\{\texttt{Outcome} = \texttt{Poor}\}$. This split holds out patients within the same multicenter cohort; it is not a held-out-hospital evaluation. Four calibrated source-only candidates use age, sex, ROSC, OHCA, shockable rhythm, and TTM; hospital is excluded from every candidate and becomes the hidden context within three prespecified disjoint hospital pairs. The public release contains candidate probabilities, outcome–hospital counts, and fixed candidate–row–context log losses, but no row–hospital links. All compatible assignments are audited at $\epsilon = .005$.

Freeze and amendment boundary. The original v1 run stopped while parsing source metadata because one sex field contained the literal token `nan`. Before any target metadata were downloaded, v1.1 froze the sole amendment: pass that token as missing to the already specified source-median imputer. Target pairs, candidates, release, threshold, seeds, four policies, gates, and the rule retaining every negative or ineligible outcome were unchanged. This is therefore a separately timestamped, source-informed amendment; only the amended target stage was sealed before target opening.

Table S13: **Frozen I-CARE transfer and truth-known proof size.** G_0 is initial public worst-compatible regret. MM is the release-only minimax choice and "aligned" is the complete-link winner. L/U are post-hoc exact holder-proof bounds. JS was the primary acquisition policy; expected-gap (EG) and selection-risk (SR) were frozen secondary policies. Query entries are mean±sample SD over 100 fixed paths; entries without ± have zero SD. ET/HGB denote extra trees and histogram gradient boosting.

Pair	N	G_0	MM→aligned	Holder L/U
A–B	175	.01669	ET→ET	18/25
E–F	50	.03705	HGB→HGB	4/4
C–D	42	ineligible: two empty outcome–hospital cells		

Pair	JS	EG	SR	Uniform
A–B	76	$57.48 \pm .64$	87	99.38 ± 10.99
E–F	17	11	14	20.47 ± 5.56

Both eligible releases begin unresolved, and all 800 policy paths stop with an exact ϵ-certificate and zero false certificates. Primary JS saves 23.5% on A–B but 17.0% on E–F, so its prespecified all-setting 20% gate does not pass. Frozen-secondary expected-gap saves 42.2% and 46.3%; this is cross-task transfer evidence, not a retrospectively promoted primary result. Its mean planning time is 1.99/.124 seconds per path versus .053/.012 seconds for uniform, explicitly trading more computation for fewer disclosed links. The release-only minimax choices already equal the aligned winners, so the decision-repair gate also does not pass. Holder-side analysis separates proof from discovery: E–F has an exact four-link proof, whereas A–B remains rigorously bracketed at 18–25 links while expected-gap needs 57.48 online links. C–D is retained as ineligible rather than replaced. An independent implementation reconstructs the released arrays, every answer, every preceding unresolved state, and every first stop from the frozen artifacts.

On this new multi-center task, nonzero disclosure remains necessary and the one-step expected-gap priority transfers descriptively, while the frozen primary and material-repair gates are allowed to fail.

S2.8 ZERO-LINK SUFFICIENCY

Protocol-frozen confirmatory audit: UCI Diabetes (zero-link boundary). The protocol-frozen UCI Diabetes audit was already resolved at zero links. On the 4,096-row target, ExtraTrees was both the deterministic release-only action and the complete-alignment winner. Its worst-compatible regret was $0.0028950183 < .005$, and its realized regret was zero. The randomized release-only

rule had worst-compatible regret 0.0019055357 and realized regret 0.0003783527. All 200 frozen JS and uniform paths stopped without paid queries; 26 identities were implied by public quotas. Both prespecified release-only true-regret gates of $.010$ failed, so the primary efficacy criterion was not met. Independent provenance, source-refitting, and stopping-rule checks passed on the unchanged artifacts. The result establishes zero-link sufficiency: no acquisition is needed under this contract. The timestamp, verification receipts, and continuation history are recorded in the reproducibility ledger below.

Protocol-frozen cross-modality audit: PTB-XL. PTB-XL supplies patient-respecting ECG folds and PTB-XL+ supplies harmonized ECGDeli features (Wagner et al., 2020; Strodthoff et al., 2023). Before the official files were present, a timestamped protocol fixed myocardial-infarction superclass prediction, age at least 65 as the hidden context, one lowest-ID ECG per patient, folds 1–8/9/10 for fitting/calibration/target, four CPU candidates, and the same $\epsilon = .005$ JS/uniform comparison. The source-only pipeline used 15,023 fitting patients, 1,942 calibration patients, and 529 ECGDeli features; the target contained 1,904 patients. Histogram boosting had exact release-only minimax regret zero and remained the complete-alignment winner with regret zero. A non-importing verifier replayed all 200 first stops, which all occurred at zero paid links with no false certificate. This is a positive cross-modality zero-link result: it validates the audit's ability to certify that disclosure is unnecessary, but supplies neither a nonzero-query policy comparison nor a repaired choice.

Protocol-frozen cross-country audit: UCI Sepsis. Before downloading the official archive, the protocol fixed Norway's 110,204-admission primary cohort as source, the independent 137-patient South Korea validation cohort as target, in-hospital death as outcome, sex as hidden context, four source-only candidates, and $\epsilon = .005$ (Chicco & Jurman, 2020). The single target opening found histogram boosting to be both the release-only minimax action and the aligned winner. An independent implementation converted every released binary64 loss to an exact rational, recomputed all 16 pairwise quota extrema, and certified $G_0 = 0$ and aligned regret zero. Hence every frozen policy stops before dispatch and no policy comparison is defined. The frozen configuration omitted a generic maximum-query field; rather than amend it after opening, a separate zero-query closure records that the cap cannot affect an already resolved release. The small validation cohort supplies a cross-country zero-link sufficiency example under the frozen-table contract.

Among the primary settings, only Heart H–S changes the deterministic minimax action materially; the other nonzero-query rows repair the guarantee rather than the selected action. On opened CIFAR, the post-hoc minimax action and the conditional-independence point completion both select DenseNet; the former is now independently reconstructed over the complete release rather than inferred from that point completion. A companion privacy-safe audit recomputes the eleven core rows from 19 SHA-256-bound source receipts, fails closed on status and bracket ordering, and stores no row identifiers or context vectors.

S2.9 Exact holder certificates and numerical verification

Why some multi-candidate frontiers close. Tables S14–S15 show the pairwise maxima and their minimum over *all* truth-feasible candidates, the global selection lower bound. On COIL, Waterbirds, and Heart, its bottleneck certificate is already safe against every other challenger, so that same set is a matching upper bound. On CelebA, the 263-row bottleneck certificate remains unsafe against challengers 0 and 3; the pairwise sets overlap imperfectly (Jaccard 0.254 and 0.358 with the bottleneck), and their union gives an initial 341-row certificate. Joint optimization restricted to that union removes 37 rows and returns an exact-checked 304-row certificate; unrestricted optimization removes two more. The resulting 302-row certificate leaves a 39-row gap between the certified lower and upper bounds.

The audit uses aggregate exact regrets, not a point-winner shortcut. The closest excluded candidates have regret $0.228456/0.242453$ on Waterbirds validation/test, 0.005579 on CelebA at .005, and 0.010491 on CelebA at .010; COIL has no excluded candidate at either displayed threshold.

Table S14: **Multi-challenger certificate structure at $\epsilon = .005$.** Pair costs are ordered by challenger index after fixing the complete-alignment selected candidate. U_{cert} is the size of a verified sufficient set; CelebA reports the tighter unrestricted exact-checked incumbent, without promoting the numerical solver bound to exactness; "bottleneck joint?" asks whether the hardest pair certificate alone protects against every challenger. Heart H/S/VA abbreviate Hungary, Switzerland, and VA Long Beach.

Target	Selected	Pair costs	Bottleneck	U_{cert}	Bottleneck joint?	Global result
COIL-2000	logistic	145/159/31	159	159	yes	exact 159
Waterbirds val	CLIP-L/14	123/41/0	123	123	yes	exact 123
Waterbirds test	CLIP-L/14	619/231/0	619	619	yes	exact 619
Heart H–S	HGB	35/86/64	86	86	yes	exact 86
Heart H–VA	HGB	0/71/39	71	71	yes	exact 71
Heart S–VA	HGB	0/69/25	69	69	yes	exact 69
CelebA	EfficientNet-B0	127/263/200	263	302	no	263–302

Table S15: **All-candidate quantifier audit.** Candidate indices follow the frozen report order. The exact-dyadic verifier first identifies every candidate with complete-alignment regret at most ϵ, reports each L_a, and only then takes $L = \min_a L_a$.

Target	ϵ	Truth-feasible candidate:L_a	L	U_{cert}	Result
COIL-2000	.005	0:159, 1:414, 2:437, 3:413	159	159	exact
COIL-2000	.010	0:0, 1:0, 2:0, 3:0	0	0	exact
Waterbirds val	.005	2:123	123	123	exact
Waterbirds val	.010	2:120	120	120	exact
Waterbirds val	.020	2:113	113	113	exact
Waterbirds test	.005	2:619	619	619	exact
Waterbirds test	.010	2:600	600	600	exact
Waterbirds test	.020	2:564	564	564	exact
Heart H–S	.005	3:86	86	86	exact
Heart H–VA	.005	3:71	71	71	exact
Heart S–VA	.005	3:69	69	69	exact
CelebA	.005	2:263	263	302	bracket
CelebA	.010	1:318, 2:174, 3:308	174	195	bracket

Endpoint and numerical contract. The frozen `loss_if_context` binary64 values are interpreted as exact dyadic rationals; ϵ is the exact decimal token in the production record. Exact contrasts are formed only after this conversion: subtracting two binary64 values first and then rationalizing the rounded difference is not an accepted certificate. Thus the reported integer minimum is decided by exact comparisons, not a solver tolerance. Table S16 shows the bottleneck frontier at the first feasible budget k and its predecessor. For the displayed bottleneck ordered pair (a, j), define $F(k) = \min_{|S|\le k} \Phi_{aj}(S)$ over disclosures using the holder-known true answers; F is local notation for this pairwise residual contrast, not the completion fiber $\mathcal{F}$. Displayed decimals are rounded; the author-side verifier records the rational numerator and denominator for every entry. The public pairwise routine uses its floating-point frontier only to locate a candidate budget, then recomputes the prefix allocation and both endpoint inequalities with exact integer arithmetic. Its tolerance therefore cannot turn $F(k-1) > \epsilon$ into a smaller "exact" certificate. The common-order joint routine and the selection bracket use the same original-loss arithmetic for truth feasibility and every returned joint witness. The empirical-lattice regret-radius routine is fail-closed: thresholds at or below its declared numerical resolution return `NUMERICALLY_UNCERTAIN`; its numerical MILP is an outward relaxation, and `OPTIMAL` is returned only when the rounded integer witness passes the exact original-loss comparison against the requested decimal threshold.

Table S16: **Exact $k-1/k$ boundary audit.** Positive columns verify $F(k) \le \epsilon < F(k-1)$ for the bottleneck pair. These certify the pairwise lower endpoint; global exactness additionally requires the joint test above.

Target	ϵ	k	$F(k)$	$\epsilon - F(k)$	$F(k-1) - \epsilon$
COIL-2000	.005	159	.004997040	2.960×10^{-6}	6.058×10^{-6}
Waterbirds val	.005	123	.004900539	9.946×10^{-5}	1.269×10^{-3}
Waterbirds val	.010	120	.009013279	9.867×10^{-4}	3.928×10^{-4}
Waterbirds val	.020	113	.018802325	1.198×10^{-3}	2.441×10^{-4}
Waterbirds test	.005	619	.004922445	7.756×10^{-5}	1.863×10^{-4}
Waterbirds test	.010	600	.009965573	3.443×10^{-5}	2.333×10^{-4}
Waterbirds test	.020	564	.019749228	2.508×10^{-4}	2.572×10^{-5}
Heart H–S	.005	86	.004681683	3.183×10^{-4}	2.414×10^{-4}
Heart H–VA	.005	71	.004563941	4.361×10^{-4}	1.961×10^{-5}
Heart S–VA	.005	69	.004777337	2.227×10^{-4}	3.372×10^{-4}
CelebA	.005	263	.004998558	1.442×10^{-6}	3.990×10^{-5}
CelebA	.010	174	.009934418	6.558×10^{-5}	9.361×10^{-6}

Construction and public verification cost. The reusable frontier is recomputed so each threshold record is self-contained. Table S17 separates pairwise/greedy holder construction from a public check receiving only authenticated answers and the release; time and memory are diagnostics, not complexity claims. On CelebA, unrestricted search took 2403.4/276.7 seconds at $\epsilon = .005/.010$, far longer than pairwise/greedy construction. Public checking took 0.081 seconds and independent audit 45.49 seconds; these are single-host phase diagnostics, not general end-to-end costs.

Table S17: **Disclosure computation audit.** Holder times exclude optional unrestricted joint-MILP search. The independent verifier imports no production module.

Target	Construct (s)	Peak (MiB)	Verify (ms)	Audit (s)
COIL-2000	29.41–37.71	103.6	62.7	56.19
Waterbirds val	0.62–0.72	78.1	28.3	1.78
Waterbirds test	14.79–15.04	76.6	136.9	40.76
CelebA	12.36–37.54	75.0–76.5	81.4	45.49

CelebA source heads use 300 iterations; calibrators stopped in 8–9. This hidden-annotation study is not external-holder validation or a measured acquisition-cost study. The anonymous ZIP omits private oracles and path transcripts. Its generic evidence aliases and code permit selected aggregate checks, but it does not ship CelebA's full public loss table or all named result ledgers; those need separate access. Changed prediction bytes require a new certificate.

S2.10 OBSERVED-ALIGNMENT AND NATIVE ADEQUACY AUDITS

Native-probability audits. ACSIncome fits five pipelines on 2017 and tests 500,000 untouched 2018 rows across 51 jurisdictions; all 11 gates pass (Ding et al., 2021). CivilComments fits five pipelines on 267,516 train/132,730 test rows while reserving nine identity indicators; the word–character union wins both axes, context certifies the top pair, and 2/18 point changes lie outside the simultaneous family (Koh et al., 2021; Borkan et al., 2019). **Pooled intervals.** ACSIncome marginal/context losses are .425466/.423229; contrasts are .0002815 $[.0001112, .0004518]$/.0001766 $[.0000059, .0003474]$. CivilComments losses are .206054/.203655; contrasts are .000798 $[-.000293, .001889]$/.001229 $[.000165, .002294]$ (simultaneous 95%).

A frozen native audit certifies decision adequacy. Cleaned DomainNet freezes three pairs and $\epsilon = .005$ on 409,832/176,743 train/test images. An exact-rational dual audit gives raw/temperature bounds $0/0$, $0.001691/0$, and $0/0$; 0/3 reversals and all six winners are certified. This is an adequacy certificate, not an initially unresolved acquisition task.

S3 OBSERVED-ALIGNMENT INFERENCE: ERROR CONTROL AND POWER

This section evaluates inference after row alignment is observed, not the exact fixed-release link certificates. Here the evaluator has aligned losses; uncertainty comes from sampling rather than hidden row–context links. The fixed-release disclosure and acquisition audits precede this section in Section S2.

The studentized multiplier mode puts every marginal pair, conditional pair, and adoption statistic in one asymptotic max-$|T|$ family. The frozen campaign has six critical error-control cells (2,000 independent worlds each) and ten designed reversal cells (500 each), totaling 17,000 displayed-grid worlds. Two retained stress cells add 200 and 100 worlds. All six procedures run on every world, so the ledger contains $6\{12{,}000 + 5{,}000 + 200 + 100\} = 103{,}800$ method outcomes. The aligned-data multiplier audit's worst critical-cell false-direction rate is 0.0050 (Wilson 95% $[0.0027, 0.0092]$), versus 0.0505 uncorrected and 0.9990 conditional-only in the hardest near tie. The bounded-loss Hoeffding mode has a finite-sample union bound under independent, almost-surely bounded units (the finite-state and clipped-t_4 cells here) and no observed false direction, but is unresolved across the critical grid. Nine of ten power cells declare INVALID in every replicate; the misspecification cell reaches 0.916. Proposition 11 states the distribution-free bounded-loss certificate.

Frozen simulation specification. The retained campaign is bound to the archived runner whose SHA-256 begins `0c229d5c4fd0`, not to the earlier full-grid generator. The executable snapshot, cell configuration, final population vectors, and one recorded repetition per cell are supplied in `examples/aligned_simulation/`. The archive also includes all 103,800 retained method outcomes and a streaming aggregate verifier. Its full-ledger recount checks final truths, repetitions, event counts, and denominators; retained pairwise error flags are not regenerated. This differs from the 18 repetition-zero sample reruns and from an optional full campaign rerun. Before world-specific perturbations, marginal and conditional values are $s\,\mathrm{linspace}(.16, .04, M)$ and $s\,\mathrm{linspace}(.14, .02, M)$ for $s \in \{1, .6, .18\}$ (large, medium, near-tie). Declared reversals swap coordinates and add .035, .025, .015, or .020; the one-beneficial cell sets one conditional value to .07. Low/high correlation use shared–idiosyncratic noise scales $.008/.025$ and $.025/.010$. Finite cells use symmetric two-point noise, Gaussian cells use normal noise, and heavy cells use clipped, variance-scaled t_4 noise. Cell names describe configurations before calibration, not the final truth; Table S18 records those configurations. More precisely, for candidate m, let m_m and v_m denote the transformed marginal and conditional population values above, and let $\delta \in \{0, .01, .025\}$ denote the none/moderate/strong calibration shift. The generator sets

$$a_m = \mathrm{linspace}(-\delta/2, \delta/2, M)_m, \qquad b_m = \mathrm{linspace}(-\delta, \delta, M)_m,$$

and replaces (m_m, v_m) by $(m_m - a_m, v_m - b_m)$. With exactly $\mathrm{round}(n\pi)$ positives (clipped to $[2, n-2]$ and shuffled), $k_i = 1.35$ for $Y_i = 1$ and $k_i = 1$ otherwise, the four row losses are

$$L_i^0 = .72 + u_i, \quad L_i^C = .55 + .8u_i, \quad L_{im}^Z = L_i^0 - m_m + k_i e_{im}, \quad L_{im}^{C,Z} = L_i^C - v_m + k_i f_{im}.$$

For low/high correlation, the scales of u_i and (e_{im}, f_{im}) are $(.008, .025)$ and $(.025, .010)$. Each noise coordinate is independently Rademacher at its scale, Gaussian $N(0, \sigma^2)$, or $\sigma\,\mathrm{clip}(T_4/\sqrt{2}, -3, 3)$

according to the DGP column. Variable clustering draws K_i uniformly from $\{1, \ldots, 5\}$ and repeats the entire sampled unit K_i times while retaining its unit identifier for cluster-level inference. The runner asserts every realized loss is in $[0, 1]$. The finite-state and clipped-t_4 noise supports imply bounded losses under these frozen parameters. Gaussian losses are not almost-surely bounded; passing the realized-loss assertion does not give them a distribution-free Hoeffding guarantee. Multiplier results remain simulation operating characteristics under the displayed DGPs. The truth transformation is exact: c01/c02 use the baseline vectors; c03/c04 set $m_2 = m_1$ and $v_2 = v_1 + .01$; c05/c06 replace the conditional vector by $-\text{linspace}(.01, .06, M)$. The margin, noise, clustering, and calibration columns below complete each cell specification; calibration subtracts (a_m, b_m) as defined above. Consequently c04 is INVALID/ADOPT, not a final tie, and c06 is VALID/ADOPT, not a no-useful-representation null. `truth_catalogue.json` lists every final vector, pair contrast, and truth status, including the two stress cells. For every cell, the recorded any-false-direction event is the union of a certified marginal or conditional sign inconsistent with its final population contrast and an incorrect ADOPT/ABSTAIN decision at threshold zero. Thus the reported critical-cell rate evaluates erroneous declarations, not rejection of one common global null. Reversal-detection power is the probability of declaring INVALID when final truth is INVALID; it does not require certifying every reversed pair.

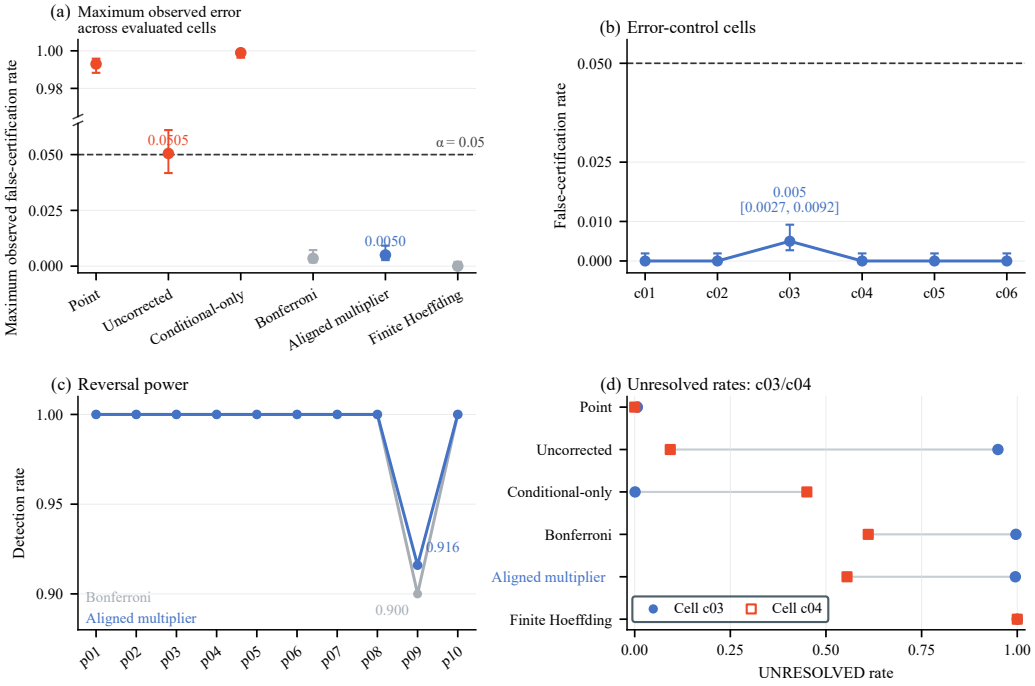

Figure S3: **Operating characteristics of the aligned-data multiplier audit.** Error control, power, and unresolved outputs are evaluated separately. Error bars are Wilson 95% intervals. "Aligned multiplier" denotes the studentized observed-alignment inference procedure, not the exact finite-table verifier. The multiplier method is grid-validated; the finite Hoeffding method is conservative and frequently unresolved. Panel (a) reports the maximum observed error across the evaluated grid, not a worst-case guarantee. Gaussian cells lack the almost-sure boundedness required by the finite-sample Hoeffding guarantee.

Table S18: **Frozen c01–c06 and p01–p10 grid.** Each c-cell has 2,000 worlds and each p-cell 500; c-cells assess error control and p-cells power. $\pi = P(Y = 1)$. "Var." means variable-size clusters. World names precede calibration; final rank/adoption truths below are exported from the supplied final population catalogue.

Cell	Configuration world	n	M	π/corr./margin
c01	valid	100	4	.5/low/medium
c02	valid	250	8	.1/high/near
c03	near tie	100	4	.5/high/medium
c04	near tie	250	8	.1/high/near
c05	none useful	100	4	.1/low/medium
c06	none useful	250	8	.1/high/medium
p01	invalid top	100	4	.5/low/large
p02	invalid top	250	4	.1/high/medium
p03	invalid top	500	8	.1/high/near
p04	invalid lower	500	16	.5/low/medium
p05	one beneficial	100	4	.1/low/medium
p06	one beneficial	500	8	.5/high/medium
p07	synergy	250	4	.5/high/medium
p08	calibration reversal	250	8	.1/high/near
p09	misspec. reversal	250	8	.5/low/medium
p10	both failures	250	8	.1/high/medium

Cell	DGP	Cluster	Calibration	Final rank/adoption
c01	finite	none	none	VALID/ADOPT
c02	Gaussian	var.	moderate	VALID/ADOPT
c03	finite	none	none	UNRESOLVED/ADOPT
c04	clipped t_4	var.	strong	INVALID/ADOPT
c05	finite	none	none	VALID/ABSTAIN
c06	clipped t_4	var.	strong	VALID/ADOPT
p01	finite	none	none	INVALID/ADOPT
p02	Gaussian	var.	moderate	INVALID/ADOPT
p03	clipped t_4	var.	strong	INVALID/ADOPT
p04	finite	none	none	INVALID/ADOPT
p05	finite	none	none	INVALID/ADOPT
p06	Gaussian	var.	moderate	INVALID/ADOPT
p07	finite	var.	none	INVALID/ADOPT
p08	Gaussian	var.	strong	INVALID/ADOPT
p09	clipped t_4	none	moderate	INVALID/ADOPT
p10	clipped t_4	var.	strong	INVALID/ADOPT

S3.1 THE GUARANTEE IS CONDITIONAL ON THE DECLARED CONTRACT

The certificate audits identifiers, grouping metadata, finite losses, and unstable standard errors, but not substantive independence, hidden clustering, or future stability. Complete declarations yield CONTRACT-COMPLETE, missing ones CONTRACT-INCOMPLETE, and future claims OUT-OF-SCOPE.

Table S19: **Inference contract for the aligned-data remedy.** Completeness means that required declarations are present, not that their substantive truth is machine-verified.

Field	Finite Hoeffding mode	Multiplier mode
Estimand	fixed $(C, \mathcal{A}, \ell, Q, \mathcal{M})$	same
Unit	independent entity/group mean	declared independent cluster
Loss	bounded in $[0, B]$	finite variance and regularity
Clipping	B prespecified; changes estimand	prespecified if used
Candidate / coordinates	frozen stack; prespecified family	same
Guarantee	finite-sample union bound	asymptotic simultaneous coverage
Not covered	hidden dependence, future-site shift, undeclared training randomness	same, plus small-sample approximation error
Failure output	CONTRACT-INCOMPLETE or UNRESOLVED	UNRESOLVED or unstable standard error

Table S20 reports a commit-frozen 600-replicate-per-cell stress protocol. Correct cluster-level analysis has a 5.5% any-false-direction rate, whereas incorrectly declaring the same clustered rows independent produces 39.8%. The bounded finite-sample mode makes no false claim here but is conservative; in the rare-catastrophe cell, increasing the prespecified clipping bound from 1 to 5 expands mean simultaneous width fivefold and changes the estimand.

Table S20: **Prespecified inference-contract stress test.** Metrics are under the null; "rows (wrong)" intentionally violates the declared independent-unit requirement. Each row uses 600 frozen replicates.

Scenario / analysis	False dir. (%)	Unresolved (%)	Mean width	Adopt FP (%)
IID bounded / finite	0.0	100.0	0.296	0.0
Clustered / cluster multiplier	5.5	94.5	0.091	1.8
Clustered / rows (wrong)	39.8	60.2	0.049	11.2
Rare loss / finite, $B = 1$	0.0	100.0	0.296	0.0
Rare loss / finite, $B = 5$	0.0	100.0	1.480	0.0

The corresponding positive-mean cell is a secondary diagnostic; the experiment does not exhaust dependence violations or establish future-domain stability.

S4 OBSERVED-ALIGNMENT DECISIONS AND SENSITIVITY ANALYSES

S4.1 SOURCE-REFIT INFERENCE CONTRACT

The end-to-end extension resamples dependency units in five disjoint roles: source-train, source-select, source-calibrate, source-evaluate, and target. Each draw refits probes, regularization, calibration, and the source action before target evaluation; target labels enter only paired evaluation, and malformed or incomplete draws fail closed. One studentized max-$|T|$ bootstrap family covers all declared stack/statistic cells. Its intervals quantify stability under this frozen resampling and fitting contract; they are not distribution-free coverage.

S4.2 FIXED-STACK SKIN DECISIONS

The PCA-16 skin analysis certifies both point-winner changes under a single 20,000-draw, 24-coordinate max-$|T|$ family: BiomedCLIP replaces DINOv2 on MSK and Handcrafted on PAD-UFES. These intervals condition on frozen source-fitted pipelines and target image-weighted risk.

Appendix H gives the training, component resampling, calibration, simultaneous intervals, and all-competitor contrasts; the source-refit analysis below has a different resampling estimand.

S4.3 SOURCE-REFIT DECISION AXES

The broader source-refit summary below uses 1,000 end-to-end bootstrap draws; the 20,000 draws above condition on one frozen source-fitted stack.

Table S21: **The three decision axes separate in aligned data.** Intervals are simultaneous over the complete declared family. "Source winner" is a bootstrap frequency, not a certified target total order.

Domain	Groups	Source winner	Rank	Adopt	Transport / adoption evidence
PAD-UFES	968	DINOv2, 889/1000	UNRESOLVED	ADOPT	UNRESOLVED; BiomedCLIP 0.311 [0.102, 0.520]
MSK	1,917	DINOv2, 886/1000	UNRESOLVED	ADOPT	UNRESOLVED; handcrafted 0.033 [0.007, 0.059], DINOv2 0.073 [0.025, 0.121]

PAD-UFES most often selects DINOv2 on source bootstrap draws, while only BiomedCLIP clears the target adoption threshold simultaneously.

S4.4 CAPACITY AND PROJECTION SENSITIVITY

Figure S4 separates the primary target–stack audits from robustness cells that reuse the same target labels. It retains the native result, all prespecified capacity points, three additional PCA-16 seeds, and a different PCA-32 probe. Both PCA-16 changes repeat across all frozen seeds; PCA-32 MLP changes both point winners but certifies neither displayed winner pair; PCA-64 logistic certifies PAD-UFES, and native winners do not change. These repeated-target sensitivity cells assess stack dependence; they are not additional independent occurrence audits.

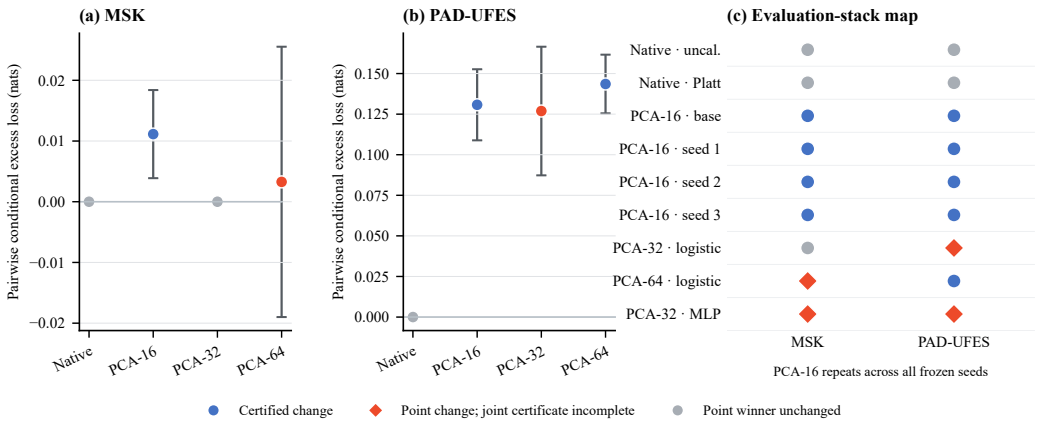

Figure S4: **Winner changes recur across PCA-16 seeds but depend on the evaluation stack.** Points and intervals show the conditional excess loss of the displayed marginal winner relative to the displayed conditional winner. Blue means that the same displayed pair has both opposite directions certified; red denotes a point change without a joint pair certificate, and gray a preserved point winner. Intervals cover both value axes and all candidate pairs within each target–stack cell, not across all sensitivity cells.

The plotted quantity is a fixed-pair contrast, not selector regret after reselection. A zero self-comparison therefore does not certify zero population selection regret. Except for the PCA-16 headline cells, the exploratory sensitivity analyses retain their original label-stratified, observed-composition resampling contracts; they are not unconditional cohort-risk claims.

The capacity map also narrows—but does not uniquely identify—the mechanism. For a fitted stack (suppressing hats and the algorithm superscript), write its standalone value as U_m, its context-assisted value as V_m, and its interaction gap as $D_m = U_m - V_m$; then $V_a - V_b = (U_a - U_b) - (D_a - D_b)$. The reported point contrasts imply interaction-gap differences of 0.0393 nats on MSK and 0.3118 nats on PAD-UFES, large enough to overcome the corresponding marginal advantages. Source-only age ladders show no positive target value from age alone. The pattern is consistent with a representation-dependent context–signal interaction under the evaluated matched-capacity stacks; these controls do not identify a unique mechanism.

A separately frozen Gaussian-random-projection control replaces PCA by source-fitted projections at dimensions 16, 32, and 64 with three fixed seeds and the same two targets. One max-$|T|$ family covers all 18 cells. Nine cells have positive simultaneous lower bounds on pairwise conditional excess loss, thereby certifying positive selection regret; eight exceed 0.005 nats. MSK has positive pairwise lower bounds in four of nine cells and PAD-UFES in five. Thus a generic low-dimensional bottleneck can reproduce the effect, so it is not uniquely PCA-specific. The seed and dimension heterogeneity, together with zero native winner changes, rules out projection-invariant or native-stack claims. These reuse the same target labels and remain mechanism-sensitivity cells rather than independent occurrence audits. The ten-row primary observed-alignment census is in Appendix H; the capacity, seed, and probe cells here reuse skin labels and do not enlarge that occurrence panel.

S5 Reproducibility and protocol chronology

The anonymous ZIP contains a sanitized code/evidence subset; detailed execution receipts and start-by-start histories remain author-side. `examples/heart_replay/README.md` gives replay commands; `manifest.json` supplies integrity hashes, not answer authentication.

Table S22: **Replay scope of the submitted anonymous ZIP.** The ZIP contains code, examples, and generic evidence aliases, but no `results/` directory. A hash identifies an omitted input, not its data.

Claim	Checkable from the submitted ZIP	Additional input for independent replay
Abstract theorems/toys	Proofs, code, and included examples.	None for the abstract statements.
CIFAR complete-joint witness	`examples/cifar_joint_replay/`: 30 cell capacities, quotas, 24 integer allocations; exact counts and 80-digit log-score replay.	None for these aggregate worlds; original inputs for image-level assignment, Brier/spherical, and TV checks.
Bank zero-link boundary	Arrays and summary aliases; not the original bank-client features.	Original rows and an alias map for full retraining.
Heart 86-link sufficiency	`examples/heart_replay/`: losses, quotas, 86 answers, independent verifier.	None; authentication is a separate governance step.
Heart 86-link minimum	`--holder` recomputes feasible candidates and pair minima 35/86/64.	Separate holder alignment is included (public UCI); public-only mode never opens it.
Heart 278-link discovery	Frozen paid transcript, quota closure, every predecessor prefix.	None for table replay; original UCI inputs for retraining.
Heart robustness	H–S 4/8/9-rule menus via `menu_sensitivity.py`; aggregate tolerance/rotation records.	H–S holder answers included; other targets and source rotations require retained original inputs.
Simulation sample rerun	`examples/aligned_simulation/verify.py`: 18 repetition-zero samples, six methods each.	NumPy/pandas; optional full campaign rerun is separate.
Simulation full recount	`verify_aggregates.py`: all 103,800 retained method outcomes, final truths, counts, denominators.	Standard-library Python; pairwise error flags are recounted, not regenerated.
Other private-context audits	Summaries and evidence aliases, not full realignment.	Original predictions, contexts, answers, and stage ledgers.

Primary-claim lookup. The main paper's complete-joint release figure is a post-hoc witness; its dataset records are indexed in `evidence/witness_index.json`. The supplied CIFAR aggregate example directly checks its 24 log-score rankings and preservation of the complete prediction law; it does not rerun optimization. The Heart entries in the main decision and disclosure tables and acquisition figure use frozen rules and policy with post-hoc minimax/quota accounting. Their common replay root is `examples/heart_replay/`: public mode checks sufficiency and first-stop prefixes, while holder mode checks the minimum. The availability table above states which inputs permit direct replay and which summaries require original data. Physical-page pointers are in the appendix guide, so source filenames alone are not treated as evidence.

Compact protocol and correction ledger. Table S23 lists ranking-artifact and acquisition changes; Heart and zero-link chronology is recorded in the Protocol chronology paragraph (p. 78). Detailed histories remain author-side.

Table S23: Protocol chronology. Post-result additions are not presented as result-blind confirmation.

Event	Timing	Scientific inputs changed?	Effect on conclusion
CIFAR exact integer certificate	Post-result	No	Converts numerical non-completion to exact feasibility; original gate retained
CXR integer-certified start selection	After frozen 119/120 run	No	Uses the same thirteen starts and gates; both ledgers remain archived
Adult converged-record selection	Post-result	No	Reports an already converged frozen start; no data or gate changed
Waterbirds baseline extension / expected-gap	Post-result	Query rule only	Completes frozen baselines and reports a post-hoc policy comparison
COIL-2000 and CelebA acquisition	Frozen before target rows were parsed	No	Both start deficient; priorities beat uniform, while the planner-superiority gate fails
DomainNet runtime/provenance amendments	Before full target inference	No	Preset gate returns 0/3; exact dual bound certifies regret at most 0.001691 nats

Waterbirds timing environment. The post-hoc audit used Linux 6.8.0-87, two 48-core Xeon Platinum 8568Y+ sockets (192 threads), 2.0 TiB RAM, Python 3.11.14, and NumPy 2.2.6. Expected-gap's break-even reveal costs against disagreement were 12.23/12.17/11.55 ms at $\epsilon = .005/.010/.020$, and against risk priority 7.83/7.70/7.37 ms: same-host timings, not hardware-independent complexity.

Protocol chronology. Heart's post-hoc quota closure retains the frozen JS trajectories but no longer charges quota-forced labels; the original accounting remains recorded. The first policy command failed before writing a trajectory because the frozen JSON omitted the full-reveal cap. A hash-bound pre-query amendment supplied only $\texttt{maximum_queries} = N$, changing no model, target, threshold, policy, seed, or gate. UCI Diabetes' verified DigiCert RFC3161 token (2026-09-21 05:06:19 UTC) preceded download. Reused-artifact verification matched source refits bit for bit, replayed 200 first stops, and recorded the randomized minimax lower witness. UCI Sepsis resolves before querying, so its omitted generic query cap cannot affect closure.