DRIFT: CAUSAL INFERENCE FOR GENERATED TREATMENTS WHEN REPLAY IS NOT PROVENANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Causal effects of generated treatments depend not only on which candidates were shown, but also on the historical mechanism that selected among them. When that mechanism is no longer reproducible, current replay estimates today's selector rather than reconstructing yesterday's assignment law. We prove that even unlimited replay and exact aggregate calibration can leave distinct historical causal effects observationally indistinguishable. This provenance boundary motivates Drift-Robust Inference From Treatment traces (DRIFT), which turns a prespecified or independently audited historical-to-current log-odds bridge into inverse-selection envelopes, observable calibration restrictions, and dual causal certificates. The resulting interval is sharp for the specified marginal trace-weight model and outer-valid for every compatible latent historical effect; at compatible zero drift it reduces to the ordered binary-pair logged counterfactual exposure functional. With bounded held-out dual scores and independent population-valid replay bands, cross-fitting yields finite-sample outer confidence sets even under approximate optimization. Regular endpoints separate historical and replay uncertainty, whereas nonregular endpoints need not admit Gaussian inference; on an explicit compact class, the minimax expected squared Hausdorff risk is of order $n^{-1} + B^{-1}$. The same dual geometry quantifies where historical audits can most reduce causal ambiguity and yields a first-order cost-aware allocation rule. A controlled HelpSteer2 study tests certificates with real ratings, constructed selectors, and potential-rating coupling; predicted inclusions are observed in valid conditions. A fixed-action real-log diagnostic shows operation under injected mismatch without validating causal coverage. Eight synthetic studies probe identification, inference, and audits: replay estimates today's selector but need not recover the historical selection law.

1 INTRODUCTION

Can historical outcomes identify the causal effect of a generated answer when only the current selection mechanism can be replayed? Consider an online tutor that generates two answers to each question and randomizes which one a student sees. A rubric fixed before outcome analysis orders eligible pairs by explanatory detail, and retention is measured one week later (Figure 1(a)). The questions, candidate pairs, displayed answers, and outcomes may survive after the retrieval state or selector version used for assignment is lost. Replay then measures the current selector, leaving the historical assignment law unknown. We call this the *mechanism-provenance problem*. It arises even for one fixed binary pair.

Generated answers rarely recur exactly: with n interactions over K_n identities and $n/K_n \not\to \infty$, per-identity replication does not diverge. Without structural sharing, the full response surface has minimax average squared error $\gtrsim \min\{1, K_n/n\}$ even under known fair randomization. A scalar randomized within-pair contrast has MSE of order n^{-1} under fixed overlap and nondegeneracy (Theorem B.1). For logged counterfactual exposures (LCE) (Athey et al., 2026), a pre-display pair and prespecified antisymmetric h define eligibility, ordering, and target $\tau_h = \mathbb{E}\{Y(1) - Y(0) \mid E = 1\}$. Local Horvitz–Thompson inference needs historical assignment probabilities (Horvitz & Thompson, 1952); structural assumptions could instead support response-surface learning.

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.

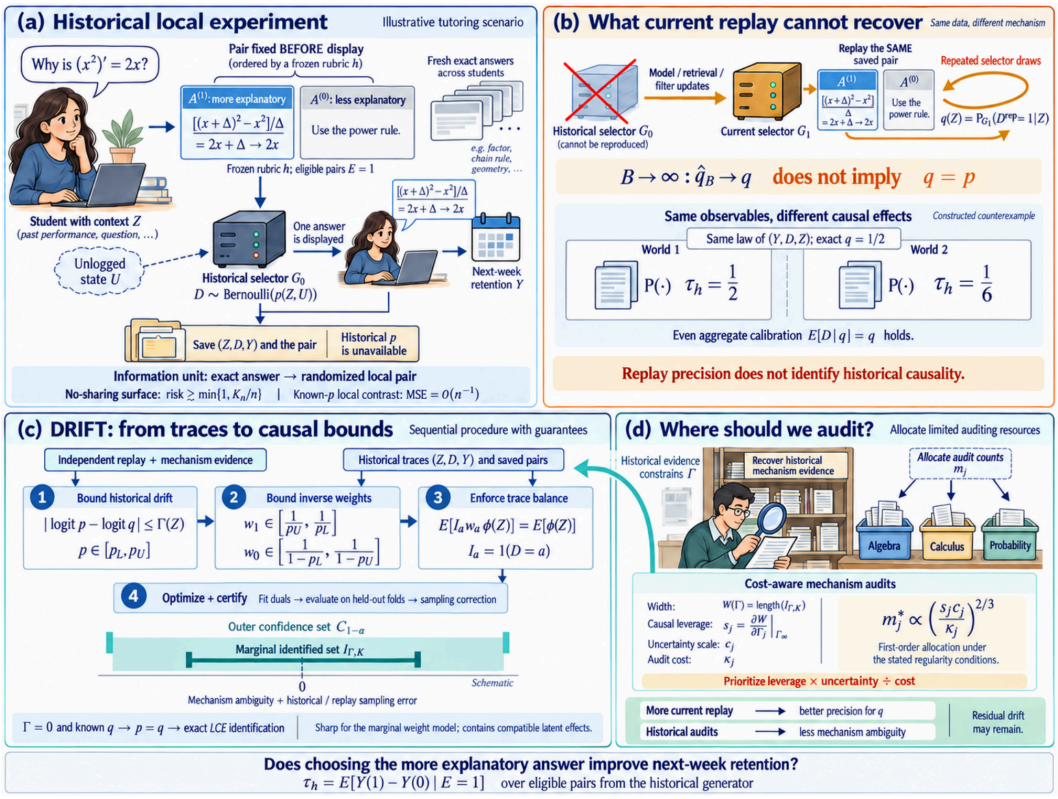

Figure 1: **Replay is not provenance.** (a) Randomized within-pair contrasts remain local targets when an unstructured exact-treatment surface cannot be learned (Theorem B.1). (b) Current replay identifies q; the two effects are a constructed counterexample (Theorem B.2). (c) A fixed or independently certified bridge and calibration yield marginal-model-sharp, latent-outer-valid sets; held-out certification needs population-valid replay bands. (d) Dual leverage s_j, uncertainty reduction c_j, and cost κ_j guide audits under the stated regularity. Intervals are schematic.

Let $p = p(Z, U)$ be the historical probability, where Z is logged context and U is lost production state, and let $q = q(Z)$ be the current selector's probability on the saved pair. Logged historical probabilities or same-mechanism replay identify p (Athey et al., 2026, Sections 9.1–9.4). Here, neither is available. As B independent replays increase, $\widehat{q}_B \to q$, but q can differ from p. Two laws can have identical historical observations, identical known replay law, and exact aggregate calibration while their causal effects equal $1/2$ and $1/6$ (Theorem B.2). They differ in the coupling of historical selection with potential outcomes.

We introduce *Drift-Robust Inference From Treatment Traces* (DRIFT). A prespecified or independently certified historical-to-current log-odds bridge bounds inverse-selection weights. Observed context moments restrict these weights, and a linear program yields an interval sharp for the specified marginal trace-weight model. Separately varying arm weights make this interval an outer bound for compatible latent effects. We adapt sensitivity-analysis geometry (Dorn & Guo, 2023; Dorn et al., 2025; Ishikawa & He, 2023; Bennett et al., 2024; Hess et al., 2026; Zhang & Zhao, 2026) to historical outcome traces and current replay. Held-out dual evaluation and population-valid replay bands give finite-sample certificates; derivatives of identified width guide historical audits (Figure 1(c)–(d)). The paper establishes:

1. **Two information boundaries.** We separate the replication barrier for an unstructured exact-treatment surface from the estimability of a scalar randomized pair contrast, then prove that unlimited current replay leaves historical effects unidentified without provenance restrictions (Theorems B.1 and B.2).

2. **Identification and two-sample inference under drift.** DRIFT combines a provenance bridge with marginal calibration, recovers LCE at compatible zero drift, and gives cross-fitted outer certificates under approximate optimization. Regular endpoints separate historical and replay errors; an explicit compact class has matching $n^{-1} + B^{-1}$ minimax squared Hausdorff risk. See Sections 4 and 5.
3. **Causal-width-directed audits.** Dual derivatives yield the first-order cost-aware allocation $m_j \propto (s_j c_j/\kappa_j)^{2/3}$. Exact-width and plug-in optimality follow under differentiability, positive leverage, and a growing consistent pilot (Section 6).

Eight synthetic diagnostics examine information location, richer restrictions, two-sample uncertainty, and finite-pilot audit allocation using archived outputs; their historical execution is not rerun in this release. The HelpSteer2 study is the primary reproducible empirical component. It uses real generated-response pairs and human-rating proxies under constructed selector and potential-rating laws, with $25{,}000$ retained per-repetition records for complete-set certification, hidden association, and sampling-versus-bridge comparisons. A fixed-action Open Bandit diagnostic uses retained outputs to study injected mismatch and simulated audits. Section 2 reviews related work. Sections 3 to 6 develop the theory; Sections 7 and 8 present experiments and implications. Proofs, extensions, and protocols follow in the appendices.

2 Related Work

LCE and generated treatments. Athey et al. (2026) study treatment-characterizing effects using an exposed item, an unexposed candidate, and their relative display probability; finite replay applies when production randomization is reproducible. We study historical assignment when replay identifies current q and historical p is unavailable. DRIFT uses calibrated partial identification, propagates replay uncertainty, and recovers their LCE functional at compatible zero drift. Compression, matrix completion, and invariant representations share information across treatment variants (Liang et al., 2026; Luo & Zhang, 2026; Su et al., 2026); smoothness can identify policy values beyond overlap (Khan et al., 2024). Our response-surface lower bound assumes no such structure.

Sensitivity analysis and partial identification. Partial identification separates implications of the observed law from those supplied by assumptions (Manski, 2003). Odds restrictions, extremal weighting, and calibration are established tools (Rosenbaum, 2002; Tan, 2006; Zhao et al., 2019); related work develops sharp quantile-balanced bounds, doubly valid inference, conditional-moment approximations, and average-case restrictions (Dorn & Guo, 2023; Dorn et al., 2025; Ishikawa & He, 2023; Zhang & Zhao, 2026). DRIFT applies this geometry with q measured from current replay. Finite-sieve balance and separately varying arm weights define its marginal relaxation. Building on inference with estimated objectives and auxiliary population information (Tudball et al., 2023), we propagate independent historical and replay samples and value new evidence by its effect on causal width.

Off-policy evaluation and provenance. Off-policy evaluation weights rewards by historical logging probabilities (Precup et al., 2000; Dudík et al., 2011; Kallus & Uehara, 2020); robust variants study transition perturbations or unobserved confounding (Bennett et al., 2024; Hess et al., 2026). DRIFT uses current replay and a provenance bridge to bound historical selection on saved pairs. We derive outer coverage and regular endpoint rates for that two-sample setting. The supplementary Open Bandit diagnostic (Saito et al., 2021) uses fixed actions and injected mismatch.

Evidence acquisition. Optimal and adaptive design allocate observations to inferential objectives (Pukelsheim, 2006; Kasy & Sautmann, 2021; Kato et al., 2024), while targeted designs acquire experiments to narrow causal bounds (Ailer et al., 2024; Maringgele & Etesami, 2026). We allocate historical-selector audits across cells. Each coefficient combines causal-width leverage, a certified Bernoulli uncertainty rate, and acquisition cost; the resulting closed form follows convex resource optimization (Rockafellar, 1970; Ben-Tal et al., 2009). Here the acquired evidence concerns selection provenance. Table 14, placed after the references, compares the assumptions and targets.

3 Information Relocation and Provenance Nonidentification

Local contrasts can remain estimable despite sparse exact-treatment recurrence (Theorem 3.1). Their estimation uses historical display probabilities. When those probabilities are lost, unlimited current replay can leave the effect unidentified (Theorem 3.2).

Theorem 3.1 (Information relocation)**.** *For outcomes in $[0,1]$ and known historical display probabilities, in the triangular-array local-pair experiment of Theorem B.1, the minimax average squared error $R_{\mathrm{surface},n}$ for an unrestricted K_n-dimensional exact-arm mean surface satisfies, for a universal $c_0 > 0$,*

$$R_{\mathrm{surface},n} \geq c_0 \min\{1, K_n/n\}.$$

In the same model, if historical display probabilities lie in $[\eta, 1-\eta]$, the local Horvitz–Thompson estimator

$$\widehat{\tau}_{\mathrm{local},n} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{D_{i,n}Y_{i,n}}{p_{i,n}} - \frac{(1-D_{i,n})Y_{i,n}}{1-p_{i,n}}\right\}$$

is unbiased for $\tau_n = \mathbb{E}(Y_{i,n}^H - Y_{i,n}^L)$ and has MSE at most $2/(n\eta)$. If the iid local scores have variance at least $v_0 > 0$, the estimator has MSE at least v_0/n. Hence $K_n \asymp n$ leaves the unstructured surface inconsistent while the scalar local contrast remains root-n estimable.

The scalar contrast uses the historical probability p. We now formalize the local experiment when p is unavailable. A binary candidate pair is fixed before display. A prespecified antisymmetric contrast $h(a,a') = -h(a',a)$ defines eligibility $E = \mathbf{1}\{h(\widetilde{A}^{(1)}, \widetilde{A}^{(0)}) \neq 0\}$ and orders eligible candidates so that $h(A^{(1)}, A^{(0)}) > 0$. Let D record the displayed candidate, Z the logged pre-display context including the pair, and U unavailable historical production state. The target $\tau_h = \mathbb{E}\{Y(1) - Y(0) \mid E = 1\}$ compares complete treatments among historical-generator pairs, excluding ties, with $Y = DY(1) + (1-D)Y(0)$. Subsequent expectations use the eligible-pair law P.

The historical choice is a genuine stochastic draw conditional on complete production state, while replay is a distinct experiment:

$$p(Z,U) = \mathbb{P}\{D = 1 \mid Z, U, Y(0), Y(1), E = 1\}, \qquad q(Z) = Q_{\mathrm{rep}}(D^{\mathrm{rep}} = 1 \mid Z, E = 1). \quad (1)$$

Assume consistency, no interference, positivity for p and q, iid eligible traces, and independent finite-stratum replay. Partitions and folds are fixed independently of history.

These conditions define a local randomized comparison. The next theorem fixes the observable law and treats q as known to isolate uncertainty about historical selection.

Theorem 3.2 (Replay-provenance nonidentification; Theorem B.2)**.** *There exist two laws satisfying* (1)*, with unobserved p, identical observable law of (Y, D, Z, q), fixed Z, known $q = 1/2$, and $\mathbb{E}(D \mid q) = q$. Historical probabilities lie in $\{1/4, 3/4\}$, but the causal effects are $1/2$ and $1/6$. No statistic based on historical observations and any number of current replay draws identifies τ_h over this class.*

The two worlds differ in how latent historical propensity couples with potential outcomes. Replay estimates q without revealing that coupling. The semi-synthetic association control below illustrates this mechanism; the theorem fixes the full observable law. We use evidence connecting historical selection to current replay for partial identification.

4 DRIFT: Calibrated Bounds and Dual Certificates

DRIFT turns a provenance bridge and calibration into a sharp marginal interval with dual certificates (Theorem 4.1; Figure 2). Compatible zero drift recovers LCE, and richer balance contracts the population interval (Corollary 4.2).

The bridge comes from mechanism evidence or an explicit sensitivity restriction. For a fixed prespecified or independently certified radius $\Gamma(Z) \geq 0$, impose

$$|\operatorname{logit} p(Z,U) - \operatorname{logit} q(Z)| \leq \Gamma(Z), \qquad p_{L/U}(z) = \operatorname{expit}\{\operatorname{logit} q(z) \mp \Gamma(z)\}. \quad (2)$$

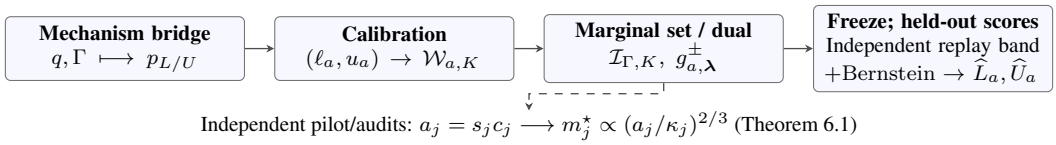

Figure 2: **DRIFT's theoretical progression.** The bridge yields a sharp marginal set (outer for latent effects) and held-out certification (Theorems 4.1 and 5.1). Compatible $\Gamma = 0$ recovers LCE; nested balance narrows population width (Corollary 4.2). Regular dual endpoints separate $V_O/n + V_R/B$ precision from r^2 marginal ambiguity; kinks limit Gaussian inference (Theorem 5.2 and Proposition 5.3). Plug-in audit optimality requires regularity and a growing consistent pilot (Theorem 6.1).

The bridge bounds historical selection. We impose observable compatibility on inverse-selection envelopes $(\ell_1, u_1) = (p_U^{-1}, p_L^{-1})$ and $(\ell_0, u_0) = ((1-p_L)^{-1}, (1-p_U)^{-1})$. For a K-dimensional basis $\boldsymbol{\phi}_K$ containing an intercept, set $I_a = \mathbf{1}(D = a)$ and $m_K = \mathbb{E}\boldsymbol{\phi}_K(Z)$. Let

$$\mathcal{W}_{a,K} = \{w \in L^\infty(P) : \ell_a(Z) \le w(Y,Z) \le u_a(Z),\ \mathbb{E}[I_a w(Y,Z)\boldsymbol{\phi}_K(Z)] = m_K\}. \tag{3}$$

Bounded outcomes and envelopes together with nonempty marginal classes give marginal sharpness. Strict interior feasibility and nondegenerate moments give strong duality and finite dual attainment (Assumptions C.2 and C.3).

A compatible latent completion projects into the marginal weights, yielding

$$\begin{aligned} \mathcal{I}_\Gamma^{\mathrm{lat}}(P,q) \subseteq \mathcal{I}_{\Gamma,K} &= [\underline{\mu}_{1,K} - \overline{\mu}_{0,K},\ \overline{\mu}_{1,K} - \underline{\mu}_{0,K}], \\ \underline{\mu}_{a,K} = \inf_{w \in \mathcal{W}_{a,K}} \mathbb{E}(I_a w Y), &\qquad \overline{\mu}_{a,K} = \sup_{w \in \mathcal{W}_{a,K}} \mathbb{E}(I_a w Y). \end{aligned} \tag{4}$$

The projection covers every compatible latent effect. Separate armwise optimization can widen the marginal set. The next result characterizes its extrema through observable scores.

For $r = Y - \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K(Z)$, let $r_+ = \max(r, 0)$ and $r_- = \min(r, 0)$, and define

$$\begin{aligned} g^-_{a,\boldsymbol{\lambda}} &= \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K + I_a\{\ell_a r_+ + u_a r_-\}, \\ g^+_{a,\boldsymbol{\lambda}} &= \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K + I_a\{u_a r_+ + \ell_a r_-\}. \end{aligned} \tag{5}$$

Theorem 4.1 (Sharp marginal identification and dual certification)**.** *Under consistency, historical conditional randomization and local p-positivity, a fixed measurable pointwise bridge, integrable potential outcomes, and a bounded basis, every compatible latent historical effect lies in $\mathcal{I}_{\Gamma,K}$ (Proposition C.4). If outcomes and envelopes are bounded and the two marginal feasible classes are nonempty, this interval is exactly the set attained by the specified marginal trace-weight model and its four arm extrema are attained (Theorem C.6). If, in addition, strict interior feasibility and the nondegenerate moment condition in Assumption C.3 hold, then for $a \in \{0,1\}$,*

$$\underline{\mu}_{a,K} = \sup_{\boldsymbol{\lambda} \in \mathbb{R}^K} \mathbb{E} g^-_{a,\boldsymbol{\lambda}}, \qquad \overline{\mu}_{a,K} = \inf_{\boldsymbol{\lambda} \in \mathbb{R}^K} \mathbb{E} g^+_{a,\boldsymbol{\lambda}}.$$

Moreover, every $\boldsymbol{\lambda} \in \mathbb{R}^K$ obeys

$$\mathbb{E} g^-_{a,\boldsymbol{\lambda}} \le \underline{\mu}_{a,K} \le \overline{\mu}_{a,K} \le \mathbb{E} g^+_{a,\boldsymbol{\lambda}}.$$

Latent projection gives outer validity; the dual reduces endpoint calculation to K dimensions. The inequalities hold for every dual vector, so numerical suboptimality can widen a certificate without reversing it. We use these scores for held-out inference and to value mechanism evidence.

Corollary 4.2 (Two nesting limits)**.** *Under the conditions of Corollary C.8, compatible zero drift collapses $\mathcal{I}_{0,K}$ to the LCE functional $\mathbb{E}[DY/q - (1-D)Y/(1-q)]$. Under boundedness, sieve completeness, and nonemptiness of every nested marginal class as in Theorem C.10, nested valid moment bases instead yield $\mathcal{I}_{\Gamma,K+1} \subseteq \mathcal{I}_{\Gamma,K}$ and Hausdorff convergence to the full-balance marginal interval.*

At compatible zero drift the weight boxes collapse. With positive drift, richer balance removes trace weights contradicted by observed moments. Evaluating a fitted certificate on the observations used to select it can narrow the resulting interval, which motivates held-out evaluation.

5 Inference with Historical and Replay Samples

DRIFT trains and freezes dual vectors before held-out score evaluation, yielding outer certificates (Figure 4 and Theorem 5.1). Two-sample rates separate precision from marginal ambiguity (Theorem 5.2). A nonregular example limits Gaussian inference (Proposition 5.3). Independently of history, a replay band must contain $q(z)$ for P_Z-almost every z with probability at least $1-\alpha_{\mathrm{rep}}$. Substituting $\widetilde{p}_L = \mathrm{expit}(\mathrm{logit}\, q_L - \Gamma)$ and $\widetilde{p}_U = \mathrm{expit}(\mathrm{logit}\, q_U + \Gamma)$ conservatively enlarges the weight boxes, defining $\widetilde{\mathcal{I}}_{\Gamma,K}$. Simultaneous binomial bands over every positive-mass homogeneous cell suffice; coverage only at realized logged contexts does not.

With $0 \le Y \le M$ and $\|\boldsymbol{\phi}_K\|_2 \le C_K$, use L fixed folds I_ℓ, $m_\ell = |I_\ell| \ge 2$, and $\pi_\ell = m_\ell/n$. Train in $\Lambda_K = \{\boldsymbol{\lambda} : \|\boldsymbol{\lambda}\|_2 \le R_K\}$ and require the replay-inflated envelopes to have a deterministic bound $\bar{w}_{\mathrm{inf}}$ for every replay realization, so every returned score obeys $|g^{\pm}| \le B_K = R_K C_K + \bar{w}_{\mathrm{inf}}(M + R_K C_K)$. This inference bound is distinct from the fixed-population envelope bound. Numerical failure invokes the prespecified zero dual vector; historical data never tune Γ. With held-out means $\bar{g}^{\pm}_{a,\ell}$ and variances $\widehat{v}^{\pm}_{a,\ell}$, define

$$\begin{aligned}
\rho_m(v, B, \delta) &= \sqrt{2v\log(2/\delta)/m} + \frac{14B\log(2/\delta)}{3(m-1)}, \\
\widehat{L}_a &= \sum_\ell \pi_\ell \{\bar{g}^{-}_{a,\ell} - \rho_{m_\ell}(\widehat{v}^{-}_{a,\ell}, B_K, \delta)\}, \\
\widehat{U}_a &= \sum_\ell \pi_\ell \{\bar{g}^{+}_{a,\ell} + \rho_{m_\ell}(\widehat{v}^{+}_{a,\ell}, B_K, \delta)\}.
\end{aligned} \tag{6}$$

The four one-sided corrections contain the arm endpoints in the directions needed for the causal interval. Combined with the population replay event, they give the finite-sample result.

Theorem 5.1 (Finite-sample outer set; Theorem D.3)**.** *Under iid eligible-history sampling with fixed folds, bounded outcomes and basis, nonempty fixed-population marginal classes, a fixed deterministic Γ, an independent population-valid replay band, and deterministic bounds on every replay-inflated envelope and returned dual score, set $\delta = (\alpha - \alpha_{\mathrm{rep}})/(4L) > 0$. Then*

$$\mathbb{P}\{\mathcal{I}_{\Gamma,K} \subseteq \widetilde{\mathcal{I}}_{\Gamma,K} \subseteq [\widehat{L}_1 - \widehat{U}_0, \widehat{U}_1 - \widehat{L}_0]\} \ge 1 - \alpha. \tag{7}$$

Every compatible latent effect satisfying the fixed pointwise bridge and the conditions of Proposition C.4 is covered; exact dual optimization and strict interiority are unnecessary.

On the replay-band event, weak duality gives the required endpoint direction, and held-out concentration accounts for sampling error. Corollary D.4 certifies known-q zero drift from collapsed boxes. Corollary D.5 extends the construction to a fixed true radius $\Gamma^\star$ and a random upper radius $\widehat{\Gamma}^U$. It allocates α_{aud} to audit failure and requires pilot, audit, and replay information to be jointly independent of the historical inference sample.

The next theorem separates endpoint precision from the ambiguity that remains with unlimited data.

Theorem 5.2 (Two-sample precision and irreducible marginal ambiguity)**.** *Under the regularity conditions of Theorem D.7, with fixed basis dimension K and replay-cell count J, the four cross-fitted arm endpoints $\widehat{\theta}$ and their population vector θ satisfy*

$$\widehat{\theta} - \theta = \frac{1}{n}\sum_{i=1}^{n} \psi_O(O_i) + \frac{1}{B}\sum_{b=1}^{B} \psi_R(R_b) + o_p(n^{-1/2} + B^{-1/2}), \qquad \mathrm{Avar}(\widehat{\theta}) = \frac{V_O}{n} + \frac{V_R}{B}.$$

If $B/n \to \kappa \in (0,\infty)$, then $\sqrt{n}(\widehat{\theta} - \theta) \Rightarrow N(0, V_O + \kappa^{-1} V_R)$. On the explicit one-cell positive-drift class of Theorem D.9,

$$c(n^{-1} \vee B^{-1}) \le \inf_{\widehat{\mathcal{I}}} \sup_{P} \mathbb{E}_P d_H^2(\widehat{\mathcal{I}}, \mathcal{I}_P) \le C(n^{-1} + B^{-1}).$$

Yet marginal point decisions retain the interval radius: their minimax risk lies between $c' \max\{n^{-1}, B^{-1}, r_{\min}^2\}$ and $C'(n^{-1} + B^{-1} + r_{\max}^2)$, and at a fixed marginal interval of radius r it is at least r^2.

The expansion separates uncertainty from historical outcomes and current replay. The minimax bound gives the corresponding two-sample rate for estimating the marginal interval. Positive-radius ambiguity remains in the point-decision bound; stronger mechanism evidence can shrink that interval.

Proposition 5.3 (Positivity does not imply Gaussian endpoints)**.** *In the positive, strictly feasible one-cell construction of Proposition D.8, the treated lower endpoint is* $4/5$*, but*

$$\sqrt{n}(\widehat{\mu}_1^- - 4/5) \Rightarrow A + |H|, \qquad A \sim N(0, 9/25), \quad H \sim N(0, 1/5), \quad A \perp H.$$

The limit is non-Gaussian, while the finite-sample outer certificate remains valid.

The active-set kink produces a non-Gaussian endpoint limit. The finite-sample certificate still applies, including to the discrete-rating experiment below. Historical audits target the remaining identification width.

6 Historical Mechanism Audits

Audits turn bridge uncertainty into an evidence-acquisition objective: historical audits constrain p, whereas replay estimates q. For $W(\mathbf{\Gamma}) = \mathrm{length}(\mathcal{I}_{\mathbf{\Gamma},K})$, the width derivative $s_j = \partial_j W(\mathbf{\Gamma})$ measures causal provenance leverage, c_j the local rate of bridge-uncertainty reduction, and κ_j acquisition cost (Proposition E.1). Together they yield first-order allocation and, with a growing consistent independent pilot, plug-in guarantees (Theorem 6.1).

In homogeneous cells, independent Bernoulli audits and replay yield simultaneous probability intervals whose opposite logit corners certify $\widehat{\Gamma}_j^U$ (Theorem E.2). Heterogeneous cells require stronger information than cell averages. The Bernoulli intervals determine how quickly each radius can be learned; the dual derivative supplies the causal coefficient $a_j = s_j c_j$ for allocating a budget.

Theorem 6.1 (Exact-width and plug-in optimal audit allocation)**.** *Let J be fixed, $\kappa_j > 0$, $C_{\mathrm{aud}} > 0$, $\Gamma_j^{\mathrm{cmp}}(m_j) = \Gamma_{j,\infty} + c_j m_j^{-1/2}$, and suppose W is Fréchet differentiable at $\mathbf{\Gamma}_\infty$ from the positive orthant, with $a_j = s_j c_j > 0$. Under the budget constraint $\sum_j \kappa_j m_j = C_{\mathrm{aud}}$, the unique continuous first-order allocation is*

$$m_j^\star = \frac{C_{\mathrm{aud}}(a_j/\kappa_j)^{2/3}}{\sum_k a_k^{2/3}\kappa_k^{1/3}}, \qquad \min_m \sum_j a_j m_j^{-1/2} = \frac{(\sum_j a_j^{2/3}\kappa_j^{1/3})^{3/2}}{\sqrt{C_{\mathrm{aud}}}}. \tag{8}$$

As $C_{\mathrm{aud}} \to \infty$, under Corollary E.4, $m^\star$ is optimal for exact marginal width up to $o(C_{\mathrm{aud}}^{-1/2})$. In the joint regime where $C_{\mathrm{aud}} \to \infty$ and an independent growing pilot consistently estimates every s_j, c_j, the plug-in exact-width excess divided by the oracle excess converges in probability to one (Theorem E.5).

Here C_{aud} is remaining audit cost, distinct from replay size B. The $2/3$ exponent comes from convex allocation; the coefficient is $s_j c_j$, which couples bridge uncertainty to the active extremal causal law. After an independent pilot, fix counts, collect fresh audit/replay data, certify the radius, and run DRIFT on reserved history (Figure 5); outcome-dependent stopping needs sequential guarantees.

Four questions organize the experiments: information location, richer restrictions, two-sample uncertainty, and finite-pilot audit allocation. EXP1–8 are archived-output diagnostics; reproducible HelpSteer2 comparisons supply mechanism and certificate evidence and fixed-radius sensitivity analysis.

7 Experiments

Eight synthetic diagnostics examine the theory under specified data-generating laws using retained outputs; their historical runner is unavailable. The HelpSteer2 (Wang et al., 2024) study uses 10,626 real prompt–response pairs with human ratings and a fixed $2{,}125/2{,}125/6{,}376$ design/development/confirmation split; historical samples draw confirmation pairs with replacement. Its potential-rating coupling and selectors are constructed (Sections F and J.2). The 20-panel main atlas combines these diagnostics and HelpSteer2 evidence (Figure 3); full diagnostics appear in the appendix.

A. Where does information live? EXP1 gives exact-surface risk 0.09504–0.11678 and roughly $n^{-1/2}$ local-contrast RMSE for the two respective estimands (Figure 3(a,b)). Compatible zero drift gives EXP2 zero width and maximum endpoint-to-LCE discrepancy 1.46×10^{-16} (Figure 11). EXP3's observationally identical worlds retain a $1/3$ truth gap as replay error vanishes (Figure 3(c,d)). Under $\mathbb{E}[p \mid Z] = q$, HelpSteer2's marginal-matching diagnostic gives replay-IPW/cell-AIPW signed errors $0.048/0.047$ with hidden rating association; without it, pointwise Monte Carlo intervals include zero (Figure 3(m)). Theorem B.2 supplies the corresponding observational-equivalence result.

B. Does richer information always tighten inference? EXP4's same-sample calibration diagnostic contracts population and implemented widths on its drift grid (Figure 3(e,f); Table 5). In EXP5, richer moments narrow population intervals but widen finite-sample outer intervals on the tested grid; the 20 infeasible systems among $6{,}400$ fold–arm–K systems invoke the declared fallback (Figure 3(g,h)). In the HelpSteer2 experiment, η denotes constructed drift strength and is distinct from the overlap bound used in the theory. Its $(\eta, \Gamma, n, B) = (0.6, 0.6, 10{,}000, 10{,}000)$ anchor yields $\mathcal{I} = [-0.070, 0.197]$ and mean certificate width 0.522; the population set and all final certificates include zero (Figure 3(n)). At the fixed observed law, assumed $\Gamma = 0.6/0.8/1.0$ gives mean certificate widths $0.522/0.635/0.744$ (Figure 3(s)). Valid balance moments contract the population interval (Corollary 4.2), but held-out certificates also pay for score estimation through B_K. EXP5 shows why population contraction need not improve finite-sample precision. Moment resolution should be chosen before evaluating held-out outcomes.

C. How do historical and replay samples determine uncertainty? EXP6's $n^{-1} + B^{-1}$ fit describes arm-order-projected endpoints and is sensitive to this projection (Section F.3.6; Figure 3(i,j)). EXP7's upper Gaussian Wald coverage is 93.2% across $1{,}000$ repetitions (Figure 11). HelpSteer2 uses two fixed folds, independent all-cell replay bands, bounded held-out scores, and the full Bernstein correction. Observed complete inclusion chains reach $1{,}000/1{,}000$ throughout 20 positive-radius conditions and two compatible zero-drift boundary checks (Corollary D.4). Point-$\widehat{q}$ breaks $\mathcal{I} \subseteq \widetilde{\mathcal{I}}$ in $854/1{,}000$ repetitions; removing Bernstein breaks $\widetilde{\mathcal{I}} \subseteq \mathcal{C}$ in $741/1{,}000$. Final $\mathcal{I} \subseteq \mathcal{C}$ fails in 0 and 45 repetitions, respectively; both cover τ_h throughout. Mean certificate widths are $1.110 \to 0.374$ as n increases from $2{,}000$ to $50{,}000$ at $B = 10{,}000$, and $0.590/0.522$ at $B = 1{,}000/10{,}000$ with $n = 10{,}000$ (Figure 3(q,r)). Panel (o) distinguishes intermediate-link violations from final set-containment failures. Panels (p) and (t) separate sampling cost from population ambiguity: larger n reduces historical-score uncertainty and larger B improves current-selector precision, but neither recovers the historical mechanism.

D. Does asymptotic audit value select the finite-pilot winner? EXP8's Full plug-in beats Leverage-only, with budget-dependent rankings against Uniform/Uncertainty-only (Figure 3(k,l)); all 10 incompatibilities among $6{,}400$ occur at the $96{,}000$ main budget, with feasibility-conditioned width means. Fixed-pilot results lack the growing-pilot allocation guarantee; HelpSteer2's bridge-radius comparisons illustrate tighter provenance restrictions.

Real-log stress test. On fixed-action Open Bandit logs, the auxiliary set $[0.00091, 0.00574]$ contains the 0.0038 randomized-log benchmark (Section F.4 and fig. 11). Injected mismatch and simulated audits of known production propensities test the auxiliary finite-action calculation. The logs lack generated pre-display pairs needed for τ_h.

8 Discussion and Conclusion

Replay estimates the current selector; provenance evidence constrains historical assignment. DRIFT combines them in calibrated sets and held-out certificates, while width derivatives define a first-order audit objective. The archived diagnostics separate population identification from finite-sample width and first-order audit value from finite-pilot rankings, and the reproducible HelpSteer2 experiment evaluates the full inclusion chain under constructed mechanisms. Its anchor certificates exclude part of the bounded outcome effect range, but every final interval contains zero. Natural platform drift, larger slates, dependence, interference, adaptive contrasts, and finite-budget audit guarantees remain open. Generated systems should archive selector versions and probabilities or audit where historical evidence most changes the causal set.

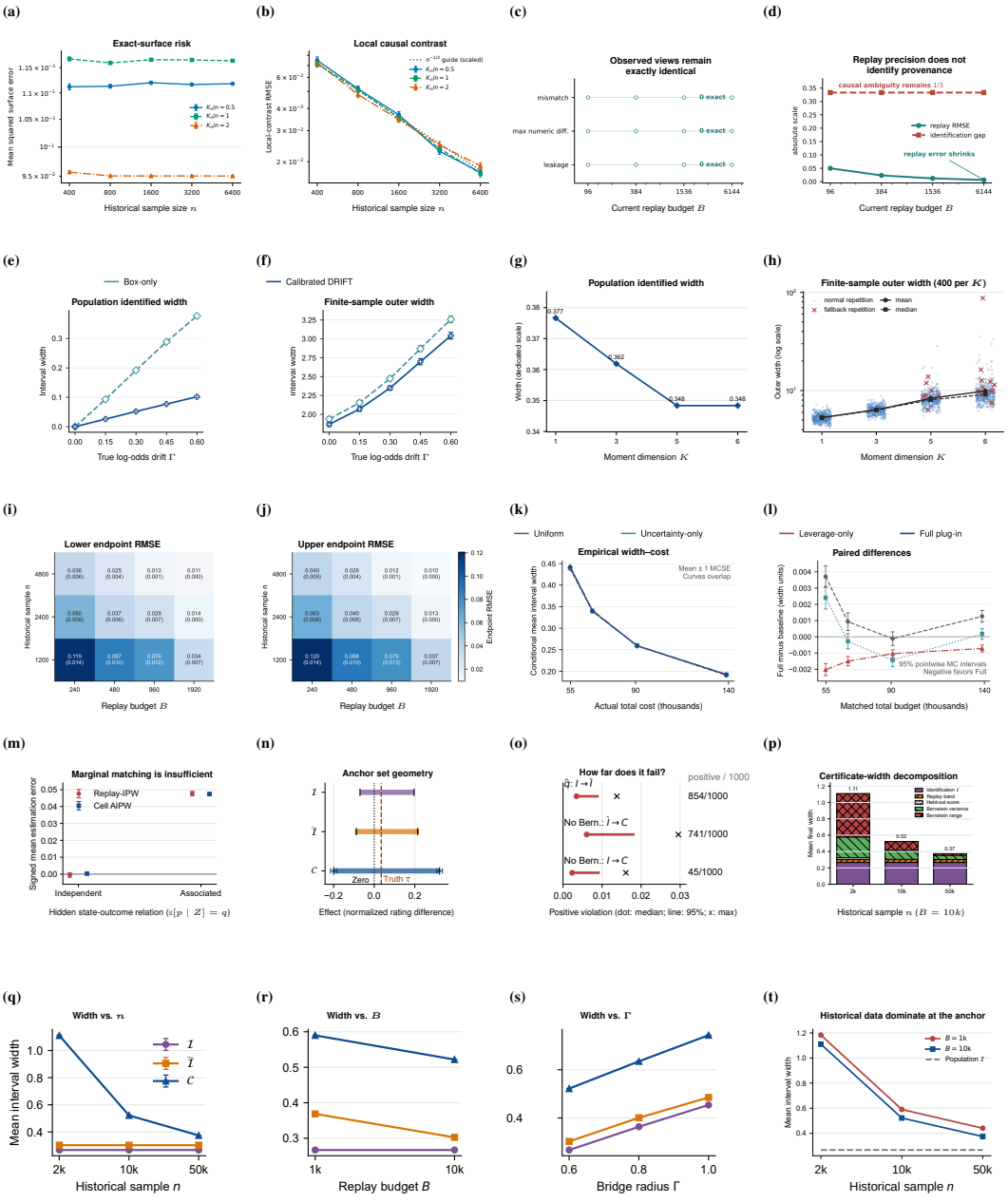

Figure 3: Four-question evidence matrix. Panels (a)–(t) run left to right across rows. Rows 1–3 preserve the original synthetic comparisons: information relocation and provenance (EXP1, 3), identification versus certificate width (EXP4, 5), and arm-order-projected endpoint fits and finite-pilot audits (EXP6, 8). Row 4 is HelpSteer2 hidden-association bias under marginal matching, anchor set geometry (horizontal scale: normalized rating difference), link-specific complete-chain violation magnitudes, and certificate-width decomposition. Row 5 shows mean widths of $\mathcal{I}$, $\widetilde{\mathcal{I}}$, and $\mathcal{C}$ against n, B, and Γ, followed by a descriptive joint $n \times B$ width comparison at $\Gamma = 0.6$, not a two-sample rate fit. These selected positive-radius helpfulness conditions fix $(\eta, K) = (0.6, 4)$ and have 1,000 repetitions each; the first three panels of row 5 show pointwise 95% Monte Carlo intervals. The first three panels of row 5 vary $n \in \{2{,}000, 10{,}000, 50{,}000\}$ at $(B, \Gamma) = (10{,}000, 0.6)$, $B \in \{1{,}000, 10{,}000\}$ at $(n, \Gamma) = (10{,}000, 0.6)$, and $\Gamma \in \{0.6, 0.8, 1.0\}$ at $n = B = 10{,}000$. The n/B contrasts use distinct, unpaired bundles; Γ shares bundles and is an assumed bridge radius. EXP2, EXP7, and other diagnostics appear in Figure 11; full panels and grids appear in Sections F and J.

AI USE STATEMENT

Generative AI tools assisted with proof–assumption cross-checks, artifact-verification code, interpretation of existing results, figure and layout inspection, and language editing. The authors manually reviewed and verified the resulting code, outputs, figures, and prose against the cited sources and released artifacts, and take responsibility for the manuscript and supplementary material.

ETHICS STATEMENT

This study combines theoretical analysis and synthetic experiments with secondary analyses of public HelpSteer2 and Open Bandit data. The released HelpSteer2 runtime table excludes prompt and response text, and the original text-bearing corpus is not redistributed. Biases in the source datasets and human ratings may limit how findings transfer across populations and settings. Our causal guarantees depend on the stated assumptions about historical selection, replay, and their relationship. The semi-synthetic results assess constructed mechanisms rather than establish causal effects in deployment. Practical applications should examine these assumptions and relevant data biases in their intended setting.

REPRODUCIBILITY STATEMENT

The companion package uses paths relative to `experiment_source/`. For HelpSteer2 it contains the frozen protocol, pre-run manifest, privacy-minimized runtime pair table, reference ratings, the $25{,}000$-row primary result table, summaries, validators, and figure-generation code. Text-bearing upstream fields are not redistributed; a pinned public revision and deterministic preparation scripts support reconstruction. For EXP1–8 and the Open Bandit analysis, retained artifacts support result inspection. `EVIDENCE_MAP.md` records the claim-to-artifact boundary, and `verify_release.py` checks manifest integrity, path safety, anonymization, the runtime corpus, and the retained HelpSteer2 result table. Further details appear in Sections J and K.

REFERENCES

Elisabeth Ailer, Niclas Dern, Jason Hartford, and Niki Kilbertus. Targeted sequential indirect experiment design. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3877. URL https://papers.nips.cc/paper_files/paper/2024/hash/dcc337bb2a4d25afefd9ab800721debb-Abstract-Conference.html.

Susan Athey, Guido Imbens, and Zoe Ji. Estimating causal effects from data generated by stochastic algorithms. *arXiv preprint arXiv:2607.05792*, 2026.

Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

Andrew Bennett, Nathan Kallus, Miruna Oprescu, Wen Sun, and Kaiwen Wang. Efficient and sharp off-policy evaluation in robust markov decision processes. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3590. URL https://papers.nips.cc/paper/2024/hash/cd1da8043ba5c1c144ab4e10a8de6e53-Abstract-Conference.html.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018. doi: 10.1111/ectj.12097.

Jacob Dorn and Kevin Guo. Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *Journal of the American Statistical Association*, 118(544):2645–2657, 2023. doi: 10.1080/01621459.2022.2069572.

Jacob Dorn, Kevin Guo, and Nathan Kallus. Doubly-valid/doubly-sharp sensitivity analysis for causal inference with unmeasured confounding. *Journal of the American Statistical Association*, 120(549):331–342, 2025. doi: 10.1080/01621459.2024.2335588. URL https://doi.org/10.1080/01621459.2024.2335588.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning*, pp. 1097–1104, 2011.

Konstantin Hess, Dennis Frauen, Valentyn Melnychuk, and Stefan Feuerriegel. Efficient and sharp off-policy learning under unobserved confounding. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=7nTKiJLkWS.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

Kei Ishikawa and Niao He. Kernel conditional moment constraints for confounding robust inference. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 650–674, 2023. URL https://proceedings.mlr.press/v206/ishikawa23a.html.

Nathan Kallus and Masatoshi Uehara. Double reinforcement learning for efficient off-policy evaluation in markov decision processes. *Journal of Machine Learning Research*, 21(167):1–63, 2020.

Maximilian Kasy and Anja Sautmann. Adaptive treatment assignment in experiments for policy choice. *Econometrica*, 89(1):113–132, 2021.

Masahiro Kato, Kyohei Okumura, Takuya Ishihara, and Toru Kitagawa. Adaptive experimental design for policy learning. *arXiv preprint arXiv:2401.03756*, 2024. URL https://arxiv.org/abs/2401.03756.

Samir Khan, Martin Saveski, and Johan Ugander. Off-policy evaluation beyond overlap: Sharp partial identification under smoothness. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 23734–23757, 2024. URL https://proceedings.mlr.press/v235/khan24b.html.

Wanting Liang, Haoang Chi, and Zhiheng Zhang. Causal representation learning with optimal compression under complex treatments. *arXiv preprint arXiv:2603.11907*, 2026.

Minrui Luo and Zhiheng Zhang. Causal matrix completion under multiple treatments via mixed synthetic nearest neighbors. *arXiv preprint arXiv:2603.11942*, 2026.

Charles F. Manski. *Partial Identification of Probability Distributions*. Springer, New York, NY, 2003. doi: 10.1007/b97478.

Tobias Maringgele and Jalal Etesami. Optimal experiments for partial causal effect identification. *arXiv preprint arXiv:2605.06993*, 2026. URL https://arxiv.org/abs/2605.06993v2. Version 2, revised September 5, 2026.

Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. In *Proceedings of the 22nd Annual Conference on Learning Theory*, 2009.

Doina Precup, Richard S. Sutton, and Satinder Singh. Eligibility traces for off-policy policy evaluation. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 759–766, 2000.

Friedrich Pukelsheim. *Optimal Design of Experiments*. SIAM, 2006.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

Paul R. Rosenbaum. *Observational Studies*. Springer, New York, 2 edition, 2002.

Yuta Saito, Shunsuke Aihara, Megumi Matsutani, and Yusuke Narita. Open bandit dataset and pipeline: Towards realistic and reproducible off-policy evaluation. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021.

Xinyan Su, Jiacan Gao, Mingyuan Ma, Xiao Xu, Xinrui Wan, Tianqi Gu, Enyun Yu, Jiecheng Guo, and Zhiheng Zhang. Orthogonal uplift learning with permutation-invariant representations for combinatorial treatments. *arXiv preprint arXiv:2602.19851*, 2026.

Zhiqiang Tan. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006. doi: 10.1198/016214506000000023.

Matthew J. Tudball, Rachael A. Hughes, Kate Tilling, Jack Bowden, and Qingyuan Zhao. Sample-constrained partial identification with application to selection bias. *Biometrika*, 110(2):485–498, 2023. doi: 10.1093/biomet/asac042. URL https://doi.org/10.1093/biomet/asac042.

Zhilin Wang, Yi Dong, Olivier Delalleau, Jiaqi Zeng, Gerald Shen, Daniel Egert, Jimmy J. Zhang, Makesh Narsimhan Sreedhar, and Oleksii Kuchaiev. HelpSteer 2: Open-Source Dataset for Training Top-Performing Reward Models. In *Advances in Neural Information Processing Systems*, 2024. doi: 10.52202/079017-0047. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/02fd91a387a6a5a5751e81b58a75af90-Abstract-Datasets_and_Benchmarks_Track.html.

Yao Zhang and Qingyuan Zhao. An average-case sensitivity analysis for unmeasured confounding. *Biometrika*, 113(2):asag030, 2026. doi: 10.1093/biomet/asag030. URL https://doi.org/10.1093/biomet/asag030.

Qingyuan Zhao, Dylan S. Small, and Bhaswar B. Bhattacharya. Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society: Series B*, 81(4):735–761, 2019. doi: 10.1111/rssb.12327.

APPENDIX CONTENTS

A Generated Treatments and Historical Local Experiments **15**

A.1 Two treatment scales: a global stream and a local experiment 15

A.2 Pair formation, historical local randomization, and current replay 16

A.3 Historical experiment and target population 17

B Two Information Boundaries **17**

B.1 Exact-arm versus local information in one experiment 17

B.2 Historical effects can remain unidentified under exact current replay 20

C DRIFT: Identification and Dual Certification **20**

C.1 Historical-to-replay bridge 21

C.2 Historical inverse-weight envelopes 21

C.3 Calibration and the marginal trace model 21

C.4 Sharp primal characterization under the marginal trace model 23

C.5 Observable dual certificates 23

C.6 LCE zero-drift corollary 24

C.7 Increasing moment resolution 24

D Statistical Inference **25**

D.1 Cross-fitting and honest dual evaluation 25

D.2 Finite-sample outer confidence sets 28

D.3 Joint historical/replay central limit theorem 29

D.4 Sampling rates and identification ambiguity 31

E Mechanism Audits and First-Order Allocation **32**

E.1 Audit value from the dual derivative 32

E.2 An auditable cell-level bridge radius 33

E.3 First-order optimal allocation 34

E.4 How the identified set determines audit value 37

F Experimental Evaluation **37**

F.1 Study design and evaluation scope 37

F.2 HelpSteer2 evidence across the four questions and boundaries 38

F.3 Eight archived-output synthetic diagnostics 41

F.4 Supplementary fixed-action Open Bandit diagnostic 47

G Discussion and Conclusion **50**

H Proofs of the Main Results **51**

H.1 Preliminary measure-theoretic facts 51

H.2 Expanding-arm and local-randomization results 51

H.3 Proof of the impossibility result 53

H.4 Validity, marginal-model sharpness, and duality 54

H.5 Unbiased replay and held-out certification 56

H.6 Asymptotic normality 58

H.7 Minimax lower bounds 59

H.8 Optimal bridge design 60

I Supporting Results and Bridge Extensions **62**

I.1 Supporting exact-arm and local statements 62

I.2 An exact-replay unbiased benchmark 64

I.3 From semantic generator drift to propensity envelopes 64

I.4 Auxiliary finite-action OPE extension 65

J Experimental Protocol, Additional Diagnostics, and Reproducibility **66**

J.1 Analytic solver for the stratified empirical program 66

J.2 Semi-synthetic mechanism-provenance protocol 67

J.3 Supplementary Experiment 4: prespecified drift-robustness protocol 70

J.4 Supplementary Experiment 5: prespecified moment-resolution protocol 71

J.5 EXP6: joint historical-replay rate protocol 71

J.6 Supplementary Experiment 7: prespecified optimized-endpoint inference protocol . 72

J.7 Supplementary Experiment 8: prespecified finite-pilot audit protocol and complete results 72

J.8 Open Bandit Dataset protocol 73

J.9 Additional numerical summaries and artifact availability 74

J.10 Reproducibility commands and boundaries 74

K Evidence Map and Four-Question Results **75**

K.1 HelpSteer2 validation across the four questions 75

K.2 Synthetic evidence across the four questions 76

K.3 Integrity and interpretation checks 76

A GENERATED TREATMENTS AND HISTORICAL LOCAL EXPERIMENTS

The observed record involves three distinct objects: the global exact-treatment stream, the eligible within-pair experiment, and current replay. We define the causal target on the within-pair scale and state the replay conditions used for population inference.

A.1 TWO TREATMENT SCALES: A GLOBAL STREAM AND A LOCAL EXPERIMENT

We observe n independent interaction snapshots indexed by $i = 1, \dots, n$. Let H_i denote pre-exposure history and A_i the exact treatment displayed at snapshot i. When it exists, the global exact-arm propensity is

$$e_n(a \mid \bar{h}) = \mathbb{P}(A_i = a \mid H_i = \bar{h}), \qquad a \in \mathcal{A}, \tag{9}$$

where $\bar{h}$ denotes a value of the history variable and is distinct from the treatment-characterizing function h defined below. The corresponding population/design support and its cardinality are

$$\mathcal{A}_n(\bar{h}) = \{a \in \mathcal{A} : e_n(a \mid \bar{h}) > 0\}, \qquad K_n(\bar{h}) = |\mathcal{A}_n(\bar{h})|. \tag{10}$$

The exact-arm lower bounds in Section B are indexed by the population support size $K_n(\bar{h})$.

At snapshot i, let $K_i(\cdot \mid H_i)$ be the historical candidate-generation kernel on finite subsets of $\mathcal{A}$. The system draws its pre-display candidate set according to

$$\mathcal{S}_i \mid H_i \sim K_i(\cdot \mid H_i). \tag{11}$$

Here K_i denotes a probability kernel, whereas $K_n(\bar{h})$ is a population arm count and the unsubscripted K used below is a calibration-basis dimension. Generation may depend on unlogged state correlated with potential outcomes; the kernel describes its history-conditional law, allowing non-random pair formation. The main analysis restricts attention to a genuinely randomized binary pair,

$$\mathcal{S}_i = \{\widetilde{A}_i^{(0)}, \widetilde{A}_i^{(1)}\}, \qquad |\mathcal{S}_i| = 2, \tag{12}$$

formed before the final display draw. The realized exact-treatment universe and its observed cardinality are

$$\mathcal{A}_{1:n} = \bigcup_{i=1}^{n} \mathcal{S}_i, \qquad \widehat{K}_n = |\mathcal{A}_{1:n}|. \tag{13}$$

The empirical occupancy statistic $\widehat{K}_n$ counts distinct exact treatments in the observed candidate stream. We use it for realized occupancy summaries and reserve $K_n(\bar{h})$ for population exact-arm results. The two quantities describe different objects.

Our causal target uses the local binary comparison rather than the global exact-arm scale. A treatment-characterizing function $h : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ is antisymmetric,

$$h(a, a') = -h(a', a), \tag{14}$$

and is fixed before analyzing the outcome. Define the eligibility, or discordance, indicator

$$E_i = \mathbf{1}\left\{h\left(\widetilde{A}_i^{(1)}, \widetilde{A}_i^{(0)}\right) \neq 0\right\}. \tag{15}$$

When $E_i = 1$, relabel the two candidates as $A_i^{(1)}$ and $A_i^{(0)}$ so that

$$h(A_i^{(1)}, A_i^{(0)}) > 0. \tag{16}$$

Candidate 1 is then the locally h-higher candidate and candidate 0 is the locally h-lower candidate. In the tutoring example, candidate 1 is the more explanatory answer according to the prespecified rubric.

Pairs with $E_i = 0$ are tied under the prespecified contrast, so the higher-versus-lower intervention is undefined for them. The target population therefore consists of eligible pairs.

For an eligible pair, the binary variable $D_i \in \{0, 1\}$ records which ordered candidate was displayed. Thus the displayed exact treatment defined above satisfies $A_i = A_i^{(D_i)}$. Let $Y_i^H = Y_i(1)$ be the

outcome that would be observed if the h-higher candidate were displayed and let $Y_i^L = Y_i(0)$ be the outcome that would be observed if the h-lower candidate were displayed. The observed outcome is

$$Y_i = D_i Y_i^H + (1 - D_i) Y_i^L. \tag{17}$$

The target is the algorithm-relative local contrast in the eligible population,

$$\tau_h = \mathbb{E}\big(Y^H - Y^L \mid E = 1\big) = \mathbb{E}\{Y(1) - Y(0) \mid E = 1\}. \tag{18}$$

The expectation averages over users, contexts, and discordant pairs generated by the historical algorithm. It is the deployment effect of choosing the h-higher rather than the h-lower candidate from that algorithm's eligible pairs. It excludes tied pairs and compares complete treatments, rather than a semantic edit that fixes all other content. Exact candidates need not recur across units.

For notational economy, all subsequent local-trace laws, expectations, and almost-sure statements are taken under the conditional distribution given $E = 1$, unless an unconditional distribution is stated explicitly. The global population support in (10) and the realized occupancy statistic in (13) remain distinct from this conditional local-trace convention.

A.2 Pair formation, historical local randomization, and current replay

The pair in (12) must be fixed before the final display draw. Conditional on this pair and the complete production state, the historical mechanism randomizes between exactly its two candidates.

Larger slates require a designated pre-display pair and its correct conditional selection law. Post-display matching to an arbitrary unshown answer cannot supply that law; the supporting finite-action calculations use a separate target.

Let $Z_i \in \mathcal{Z}$ collect logged pre-display information needed to describe the eligible local comparison: user and query features, the ordered candidate pair, pair-formation metadata, model and prompt metadata, sampler settings, and logged environment state. Let U_i denote historical production state unavailable to the analyst, such as a transient retrieval shard, unlogged moderation score, latency state, or version-specific hidden variable. The historical local selection probability is

$$p_i = p(Z_i, U_i) = \mathbb{P}\{D_i = 1 \mid Z_i, U_i, Y_i(0), Y_i(1), E_i = 1\}. \tag{19}$$

Because Z_i contains the fixed ordered pair, p_i is the probability of displaying the h-higher candidate within that pair. It describes selection within the fixed binary pair, separately from global exact-arm assignment or selection from a larger slate. The equality permits (Z_i, U_i) to be associated with both potential outcomes and the generated pair. It requires only that, conditional on the complete production state, the final choice between the two candidates is a genuine stochastic draw.

The analyst can replay a currently available mechanism on the same recorded or reconstructed ordered pair. Its probability of selecting the h-higher candidate is

$$q_i = q(Z_i) = Q_{\mathrm{rep}}(D_i^{\mathrm{rep}} = 1 \mid Z_i, E_i = 1). \tag{20}$$

These three probabilities refer to different experiments. Equation (9) concerns exact treatment identities across the population; (19) is the within-pair probability that generated the historical outcome; and (20) is the current within-pair replay probability. DRIFT uses a bridge from the third to bound the second.

To make replay a genuine second sample, separate the historical outcome traces from the current replay draws. The historical sample is

$$\mathcal{D}_n = \{O_i = (Y_i, D_i, Z_i) : i = 1, \ldots, n\}, \tag{21}$$

and the current replay sample is

$$\mathcal{R}_B = \{R_{jb} : j = 1, \ldots, J,\ b = 1, \ldots, B_j\}, \qquad B = \sum_{j=1}^{J} B_j. \tag{22}$$

Here j indexes a finite partition of eligible-pair contexts and $R_{jb} = 1$ records selection of the h-higher candidate in replay draw b from stratum j. The finite-stratum replay design is

$$R_{jb} \sim \mathrm{Bernoulli}(q_j), \qquad \mathcal{D}_n \perp \mathcal{R}_B, \qquad B_j/B \longrightarrow \rho_j \in (0, 1], \quad \sum_j \rho_j = 1. \tag{23}$$

The historical observations are iid from the eligible-pair law P. The finite partition and fold assignment are fixed independently of these observations. Replay draws are independent across draws conditional on their stratum probabilities; $\rho_1 = 1$ is allowed when $J = 1$. Thus $\mathcal{D}_n$ carries the historical outcomes, while $\mathcal{R}_B$ carries the sampling uncertainty in the currently replayable mechanism.

For finite-sample certification, replay is often run directly on the logged contexts rather than on an independent draw of contexts. In that design define

$$q_i = q(Z_i), \qquad q_{L,i} = q_L(Z_i), \qquad q_{U,i} = q_U(Z_i), \qquad i = 1, \dots, n, \tag{24}$$

and the simultaneous event

$$\mathcal{E}_{\text{rep}} = \bigcap_{i=1}^{n} \{q_i \in [q_{L,i}, q_{U,i}]\}. \tag{25}$$

The relevant guarantee is conditional-on-context simultaneous coverage,

$$\mathbb{P}(\mathcal{E}_{\text{rep}} \mid Z_{1:n}) \geq 1 - \alpha_{\text{rep}} \qquad \text{almost surely.} \tag{26}$$

This condition permits the replay draws and the logged contexts to be coupled; it only requires the replay randomization used for the band to be conditionally independent of the logged treatment and outcome draws given those contexts. This guarantee covers the realized contexts. The population confidence set in Theorem D.3 requires an almost-everywhere envelope for the population law. The independent finite-stratum design in (23), with $q(z) = q_j$ throughout each prespecified cell, supports the population replay-band condition in Section D and the two-sample asymptotic result.

A.3 Historical experiment and target population

Assumption A.1 (Pre-outcome contrast, eligibility, and consistency)**.** The antisymmetric contrast h, the eligibility rule in (15), and the pair-ordering rule are fixed without using the analysis outcomes. Eligibility is determined from the candidate pair before display. For each eligible unit, $Y = DY^H + (1 - D)Y^L$, and the displayed candidate for one unit does not affect another unit's outcome.

Assumption A.2 (Historical conditional randomization and local positivity)**.** For the eligible, pre-display ordered binary pair, Equation (19) holds for a measurable $p(Z, U)$, and there is $\eta > 0$ such that $\eta \leq p(Z, U) \leq 1 - \eta$ almost surely under $P(\cdot \mid E = 1)$.

Assumption A.3 (Replay observability and replay positivity)**.** The current replay mechanism is evaluated on the same eligible ordered pair. Its score $q(Z)$ is measurable from logged context and satisfies $\eta_q \leq q(Z) \leq 1 - \eta_q$ almost surely under $P(\cdot \mid E = 1)$. Replay randomization is conditionally independent of the historical treatment and outcome draws given its prescribed inputs. Population finite-sample inference additionally requires the independent, population-valid band in (63); the weaker logged-context guarantee (26) is distinguished explicitly.

The target is an eligible historical pair. Historical traces and current replay provide the two sources of sampling variation studied below.

From here onward, P denotes the observed law of $O = (Y, D, Z)$ conditional on $E = 1$, and unqualified expectations use this eligible-pair law. We relabel the eligible observations as $i = 1, \dots, n$; thus n in local inference is the number of eligible historical traces. The fixed contrast, pre-display pair, historical probability, and replay probability are separate objects even when they share logged metadata.

B Two Information Boundaries

When K_n/n does not vanish, expanding exact-treatment support limits recovery of an unstructured response surface. A scalar contrast instead uses randomized comparisons within pairs. If the historical selection probability is lost, even one fixed eligible pair has an unidentified effect. These are the two information boundaries behind DRIFT.

B.1 Exact-arm versus local information in one experiment

The Information Relocation Theorem places both targets in the same experiment. The response surface depends on replication across exact identities, whereas the scalar local contrast depends on randomized candidate comparisons.

Table 1: Assumptions, their roles, and empirical or design checks. Identifying conditions also require substantive support from the study design.

Condition	Role	Diagnostic or falsification route	Response to a violation
Pre-outcome contrast, consistency, no interference (Assumption A.1)	Defines the intervention and links potential to observed outcomes	Timestamp rubric and pair construction; inspect cross-unit exposure channels	Redefine the target before outcome access; use interference-aware or cluster-level potential outcomes
Historical randomization and positivity (Assumption A.2)	Makes the local contrast causal and keeps inverse weights finite	Audit randomization code and overrides; examine propensity tails where logged	Restrict to an overlap population or report broader design-based bounds; deterministic display does not support this target
Replay observability and positivity (Assumption A.3)	Identifies the current anchor q and its sampling law	Version hashes, repeated-query frequencies, cellwise coverage and support checks	Use external replay bands or a sensitivity-only analysis; do not substitute an unmatched replay score for p
Certified odds drift (Assumption C.1)	Connects current replay to the lost historical mechanism	Archived propensities, overlap traffic, shadow deployment, or prospective audits	Report a radius-indexed sensitivity curve or context-specific envelopes rather than a single identified set
Bounded outcomes and envelopes (Assumption C.2)	Gives compact fixed-population classes and controls scores; replay-inflated inference additionally needs its own deterministic envelope bound	Check outcome scale, clipping policy, replay overlap, and both fixed and inflated maximum implied weights	Use tail-adaptive concentration or truncation with an explicit bias term; the present finite-sample theorem no longer applies verbatim
Compatibility and interiority (Assumption C.3)	Supports strong duality, dual attainment/coercivity, and stable finite optimizers; primal compactness needs only nonemptiness plus bounded envelopes	Primal feasibility, slack-to-box diagnostics, redundant moments, and Hessian conditioning	For a separately specified design, drop incompatible moments; for the present fixed design, do not use the same historical outcomes to alter Γ or the moment basis, and instead use a prespecified weak-dual certificate such as the zero vector
Sieve completeness (Assumption C.9)	Justifies the $K \to \infty$ full-balance limit	Out-of-sample balance residuals across nested bases	Report each finite-K interval under its own moment restrictions
Endpoint stability (Assumption D.6)	Supports the joint Gaussian expansion, not identification	Optimizer perturbations, residual mass near kinks, and Hessian eigenvalues	Use the finite-sample outer certificate or nonregular inference rather than Gaussian endpoints

Table 2: Notation for the generated-treatment experiment.

Symbol	Meaning
$K_i(\cdot \mid H_i)$	Historical kernel generating the pre-display candidate set
$\mathcal{S}_i$	Generated binary candidate pair fixed before display
h	Prespecified antisymmetric treatment-characterizing function
E_i	Eligibility: the pair is discordant under h
D_i	Indicator that the ordered h-higher candidate was displayed
$p_i = p(Z_i, U_i)$	Historical selection probability within that ordered pair
$q_i = q(Z_i)$	Current replay probability on the same ordered pair
τ_h	Mean higher-minus-lower outcome over historical eligible pairs
$K_n(\bar{h}), \widehat{K}_n$	Population exact-arm support size and observed candidate-stream occupancy
K	Calibration-basis dimension introduced in Section C
B	Total replay draws in inference
C_{aud}	Remaining historical-audit cost budget in Section E
B_j	Number of current replay draws in cell j

Theorem B.1 (Information Relocation Theorem)**.** *Fix $\eta \in (0, 1/2]$. For each n, let $K_n \geq 2$ and let $\mathcal{T}_n = \{1, \ldots, K_n\}$ be the exact-treatment universe. For $i = 1, \ldots, n$, generate iid units (for each fixed n): an ordered pair $(A_{i,n}^{(0)}, A_{i,n}^{(1)})$ uniformly without replacement from $\mathcal{T}_n^2$, oriented by*

a fixed prespecified antisymmetric treatment contrast as $(A_{i,n}^H, A_{i,n}^L)$*. Assume this contrast has no ties on distinct treatments in* $\mathcal{T}_n$*. Let* $Y_{i,n}(a)$ *be the potential outcome of exact treatment* a*, assume* $0 \le Y_{i,n}(a) \le 1$*, and write*

$$Y_{i,n}^H = Y_{i,n}(A_{i,n}^H), \qquad Y_{i,n}^L = Y_{i,n}(A_{i,n}^L).$$

Conditional on the complete pre-display state

$$\mathcal{G}_{i,n} = \sigma\Big(A_{i,n}^{(0)}, A_{i,n}^{(1)}, \{Y_{i,n}(a) : a \in \mathcal{T}_n\}, p_{i,n}\Big),$$

draw

$$D_{i,n} \mid \mathcal{G}_{i,n} \sim \text{Bernoulli}(p_{i,n}), \qquad \eta \le p_{i,n} \le 1-\eta,$$

and observe $A_{i,n} = A_{i,n}^H$ *and* $Y_{i,n} = Y_{i,n}^H$ *when* $D_{i,n} = 1$*, and* $A_{i,n} = A_{i,n}^L$ *and* $Y_{i,n} = Y_{i,n}^L$ *when* $D_{i,n} = 0$*. Let* $\mathcal{O}_n$ *denote the resulting trace, including both candidates,* $p_{i,n}$*,* $D_{i,n}$*, and* $Y_{i,n}$*. Then the following two statements hold in this same model class.*

Let $\mathfrak{P}_n$ *denote the class of all triangular-array laws satisfying the construction above, with bounded potential outcomes and no structural relation imposed across exact-treatment identities. The exact-arm statement below is a minimax lower bound obtained by restricting this class to the displayed Bernoulli submodel; the local statement holds uniformly over the full class.*

(i) ***Exact-arm surface.*** *Consider the no-sharing submodel in which, for an unrestricted vector* $\mu_n = (\mu_{1,n}, \dots, \mu_{K_n,n}) \in [1/4, 3/4]^{K_n}$*,*

$$Y_{i,n}(a) \overset{\text{ind}}{\sim} \text{Bernoulli}(\mu_{a,n}), \qquad p_{i,n} \equiv \frac{1}{2}.$$

Here the potential-outcome vector is independent of the generated pair. With

$$R_{\text{surface},n} = \inf_{\widehat{\mu}_n(\mathcal{O}_n)} \sup_{\mu_n \in [1/4,3/4]^{K_n}} \mathbb{E}_{\mu_n}\left[\frac{1}{K_n}\sum_{a=1}^{K_n}(\widehat{\mu}_{a,n} - \mu_{a,n})^2\right],$$

there is a universal constant $c_0 > 0$ *such that*

$$R_{\text{surface},n} \ge c_0 \min\left\{1, \frac{K_n}{n}\right\}.$$

(ii) ***Local algorithm-relative contrast.*** *Define*

$$\tau_n = \mathbb{E}\left(Y_{i,n}^H - Y_{i,n}^L\right), \qquad \widehat{\tau}_{\text{local},n} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{D_{i,n}Y_{i,n}}{p_{i,n}} - \frac{(1-D_{i,n})Y_{i,n}}{1-p_{i,n}}\right].$$

Then $\widehat{\tau}_{\text{local},n}$ *is unbiased for* τ_n *and*

$$\mathbb{E}\left[(\widehat{\tau}_{\text{local},n} - \tau_n)^2\right] \le \frac{2}{n\eta} \le \frac{2}{n\eta^2}.$$

If, in addition, the common variance of the iid local Horvitz–Thompson scores is bounded below by a constant $v_0 > 0$*, then the estimator has MSE at least* v_0/n *and hence is of order* n^{-1}*.*

Consequently, whenever $K_n \asymp n$ *and the overlap and nondegeneracy constants are fixed,*

$$R_{\text{surface},n} \gtrsim 1, \qquad \mathbb{E}\{(\widehat{\tau}_{\text{local},n} - \tau_n)^2\} \asymp n^{-1}.$$

The scalar estimator pools randomized candidate comparisons, while surface estimation requires observations for each exact arm. Part (i) assumes no structural sharing across the K_n *treatments.*

The lower bound applies to an unstructured response surface. A randomized local comparison still estimates the scalar contrast. Appendix I.1 collects supporting overlap, exact-arm, and local-identification results. The next result considers the historical selection probability required for that comparison.

B.2 HISTORICAL EFFECTS CAN REMAIN UNIDENTIFIED UNDER EXACT CURRENT REPLAY

All causal effects and probability statements in the remainder of this section concern the eligible ordered-pair population; tied pairs are outside the target by construction.

The Information Relocation Theorem establishes scalar estimation under known historical display probabilities. The next theorem fixes the entire observable law and shows that current replay cannot recover a missing historical selector.

Theorem B.2 (Replay-provenance nonidentification)**.** *There exist two full-data laws $P^{(1)}$ and $P^{(2)}$ satisfying Assumptions A.1–A.3 on the eligible population with the following properties:*

(i) the logged context Z is fixed and common to the two laws, and the observable law of (Y, D, Z, q) is identical under $P^{(1)}$ and $P^{(2)}$;

(ii) $q \equiv 1/2$ is known exactly, and $\mathbb{E}(D \mid q) = q$ under both laws;

(iii) the historical probabilities take values in $\{1/4, 3/4\}$, so both laws satisfy (28) for every $\Gamma \geq \log 3$;

(iv) the causal effects are $\tau_h(P^{(1)}) = 1/2$ and $\tau_h(P^{(2)}) = 1/6$.

Consequently, no statistic based on the observed interactions and any number of current replay draws identifies τ_h over this class. For every estimator $\widehat{\tau}$,

$$\max_{j\in\{1,2\}} \mathbb{E}_{P^{(j)}} \left|\widehat{\tau} - \tau_h(P^{(j)})\right| \geq \frac{1}{6}, \qquad \max_{j\in\{1,2\}} \mathbb{E}_{P^{(j)}} \{\widehat{\tau} - \tau_h(P^{(j)})\}^2 \geq \frac{1}{36}. \tag{27}$$

Two observationally equivalent worlds. Consider one eligible ordered binary pair and let U be an equally likely latent production state. Candidate 1 is displayed with probability $1/4$ when $U = 0$ and $3/4$ when $U = 1$; candidate 0 always has outcome zero. In World 1, candidate 1 works exactly for the low-display state. In World 2, it never works for the low-display state and works with probability $1/3$ for the high-display state. Selection hides the difference: in either world the probabilities of observing $(D, Y) = (1, 1), (1, 0), (0, 0)$ are $1/8, 3/8, 1/2$.

Implication for replay. Replay identifies the current stochastic law q. It does not identify the historical law p or the association between historical state and potential outcomes. Aggregate calibration checks only $\mathbb{E}D = \mathbb{E}q$ and cannot reveal offsetting latent-state distortions. More replay reduces Monte Carlo error for the current mechanism while leaving historical provenance unresolved.

Role of a valid bridge. When $p = q$ conditional on production state and the eligible ordered pair, the two worlds are excluded and Theorem I.3 applies. More generally, a finite bridge, conditional calibration moments, and targeted historical audits can narrow the causal set. These links supply information about historical selection that current replay alone lacks.

Appendix I.1 gives the corresponding honest-interval bound. The next section uses a mechanism bridge to bound the effect when historical selection is unavailable.

C DRIFT: IDENTIFICATION AND DUAL CERTIFICATION

The framework has five steps:

$$\text{bridge} \longrightarrow \text{weight envelope} \longrightarrow \text{calibration} \longrightarrow \text{primal} \longrightarrow \text{dual}.$$

Throughout, P is the eligible-pair law and both selectors act on the same ordered pair. Mechanism evidence or a sensitivity restriction fixed independently of outcomes supplies the bridge. The bridge bounds historical selection; the envelope and calibration restrict inverse weights; the primal defines the marginal causal set; and the dual certifies its endpoints. This marginal model contains every compatible latent historical effect.

C.1 HISTORICAL-TO-REPLAY BRIDGE

For a context-dependent radius $\Gamma : \mathcal{Z} \to [0, \infty)$, impose

$$|\operatorname{logit} p(Z, U) - \operatorname{logit} q(Z)| \le \Gamma(Z) \qquad \text{almost surely under } P(\cdot \mid E = 1). \tag{28}$$

Equivalently,

$$p_L(z) = \operatorname{expit}\{\operatorname{logit} q(z) - \Gamma(z)\} \le p(z, u) \le p_U(z) = \operatorname{expit}\{\operatorname{logit} q(z) + \Gamma(z)\}. \tag{29}$$

If $\Lambda(z) = \exp\{\Gamma(z)\}$, historical local display odds may differ from replay odds by at most the factor $\Lambda(z)$. The zero-radius case is exact reproducibility. A scalar radius yields a sensitivity curve; a context-dependent radius can be certified from archived propensities, overlap traffic, randomized shadow deployment, configuration changes, or bridge audits.

Observed historical propensity versus replay anchor. Under the historical local-randomization design, define

$$e_P(z) := P(D = 1 \mid Z = z, E = 1) = \mathbb{E}\{p(Z, U) \mid Z = z, E = 1\}. \tag{30}$$

An observational marginal sensitivity model typically anchors latent selection odds at $e_P(z)$. Here $q(z)$ is measured by a separate current-replay experiment and need not equal $e_P(z)$. Even equality of these two probabilities does not imply $q = p(Z, U)$: in Theorem B.2, both worlds have $q = e_P = 1/2$ but different causal effects.

For known fixed q and Γ, (29) and the weight program below have the familiar box-constrained, calibrated sensitivity geometry (Tan, 2006; Dorn & Guo, 2023; Ishikawa & He, 2023). When $q = e_P$, the odds envelope is the corresponding marginal-sensitivity envelope with odds factor $\Lambda = e^{\Gamma}$. Finite-sieve calibration and independently varying arm weights define a marginal relaxation; a common latent completion adds compatibility constraints. The separate replay sample and auditable historical/current bridge determine the inference and evidence-acquisition questions studied next. The bridge governs historical plausibility, and calibration governs observable compatibility.

Assumption C.1 (Certified local odds drift)**.** Within the eligible ordered-pair population, the historical and replay probabilities obey (28) for a fixed measurable radius $\Gamma(Z)$ that is prespecified or certified independently of the historical outcomes used below.

All population identification results treat Γ as fixed. For the separately audited extension, write $\Gamma^\star$ for a fixed true cellwise drift radius and $\widehat{\Gamma}^U$ for its random audited upper bound. The audit event is

$$\mathcal{E}_{\text{aud}} = \left\{\Gamma^\star(Z) \le \widehat{\Gamma}^U(Z) \text{ for } P_Z\text{-almost every } Z\right\}. \tag{31}$$

Assumption C.1 concerns fixed radii. The finite-sample corollary in Section D incorporates random audited radii through their failure probability and independence requirements.

Generator-level semantic neighborhoods can be mapped into local probability envelopes; Appendix I.3 gives an OT construction.

C.2 HISTORICAL INVERSE-WEIGHT ENVELOPES

The induced inverse-weight envelopes are

$$\ell_1(z) = p_U(z)^{-1}, \qquad u_1(z) = p_L(z)^{-1}, \tag{32}$$

$$\ell_0(z) = \{1 - p_L(z)\}^{-1}, \qquad u_0(z) = \{1 - p_U(z)\}^{-1}. \tag{33}$$

The notation guarantees $\ell_a \le u_a$ for both local arms.

C.3 CALIBRATION AND THE MARGINAL TRACE MODEL

Assumption C.2 (Bounded outcomes, envelopes, and nondegenerate sieve)**.** For a known $M < \infty$, $0 \le Y(d) \le M$ almost surely for $d \in \{0, 1\}$ in the eligible population. The nested basis $\boldsymbol{\phi}_K : \mathcal{Z} \to \mathbb{R}^K$ contains an intercept, satisfies $\|\boldsymbol{\phi}_K(Z)\|_2 \le C_K$ almost surely, and has essentially linearly independent coordinates. The inverse-weight envelopes satisfy $1 \le \ell_a(Z) \le u_a(Z) \le \bar{w} < \infty$ almost surely for both arms, for a fixed finite $\bar{w}$. Replay positivity together with a uniformly bounded drift radius is sufficient for this envelope condition; historical positivity alone does not bound a wider analyst-specified envelope.

For $a \in \{0,1\}$ define $I_a = \mathbf{1}(D = a)$ and $m_K = \mathbb{E}\boldsymbol{\phi}_K(Z)$ under the eligible-pair law P. Let P_n denote the empirical measure of eligible observations. For a scalar r, write $r_+ = \max(r, 0)$ and $r_- = \min(r, 0)$. The feasible inverse-selection weights are

$$\mathcal{W}_{a,K} = \{w \in L^\infty(P) : \ell_a(Z) \le w(Y,Z) \le u_a(Z), \quad \mathbb{E}\{I_a w(Y,Z)\boldsymbol{\phi}_K(Z)\} = m_K\}. \tag{34}$$

Assumption C.3 (Compatibility and interiority)**.** For each $a \in \{0,1\}$ and sieve size K, there is a measurable w_a° and $\delta_K > 0$ such that

$$\ell_a(Z) + \delta_K \le w_a^\circ(Y,Z) \le u_a(Z) - \delta_K, \qquad \mathbb{E}\{\mathbf{1}(D=a) w_a^\circ(Y,Z)\boldsymbol{\phi}_K(Z)\} = \mathbb{E}\{\boldsymbol{\phi}_K(Z)\}, \tag{35}$$

where the expectations are under the eligible-pair law. Moreover, $\mathbb{E}\{I_a \boldsymbol{\phi}_K(Z)\boldsymbol{\phi}_K(Z)^\top\}$ is positive definite after removing redundant coordinates.

For $a \in \{0,1\}$, define the latent inverse propensity for the eligible ordered pair

$$\omega_1^{\mathrm{lat}} = p(Z,U)^{-1}, \qquad \omega_0^{\mathrm{lat}} = \{1 - p(Z,U)\}^{-1}. \tag{36}$$

These quantities are not observed because U and p are not logged. Their conditional projections onto the observed eligible trace are

$$w_a^\star(y,z) = \mathbb{E}\{\omega_a^{\mathrm{lat}} \mid D = a, Y = y, Z = z, E = 1\}. \tag{37}$$

A compatible latent historical completion is a full-data law $\widetilde{P}$ for $(Z, U, Y(0), Y(1), D)$ that induces the observed eligible-trace law P of (Y, D, Z), uses one common historical state U and one valid pair-level probability $p(Z,U)$ for both potential outcomes, satisfies consistency and historical conditional randomization with local p-positivity, and obeys the fixed pointwise bridge (28) to the measurable replay anchor q. Replay-band sampling and independence conditions enter the subsequent inference step. Define its causal effect by

$$\tau_h^{\mathrm{lat}}(\widetilde{P}) = \mathbb{E}_{\widetilde{P}}\{Y(1) - Y(0)\}, \tag{38}$$

and let $\mathcal{M}_{\mathrm{latent}}(P, q, \Gamma)$ be the set of arm-mean pairs

$$\mathcal{M}_{\mathrm{latent}}(P,q,\Gamma) = \left\{ \left(\mathbb{E}_{\widetilde{P}} Y(1), \mathbb{E}_{\widetilde{P}} Y(0)\right) : \widetilde{P} \text{ is a compatible latent historical completion} \right\}. \tag{39}$$

The next proposition projects each admissible latent completion onto arm-specific observed-trace weights, preserving balance and the causal mean.

Proposition C.4 (Latent drift induces feasible marginal trace weights)**.** *Suppose Assumptions A.1 and A.2 hold on the eligible ordered-pair population. Let measurable $q(Z) \in (0,1)$ and a fixed measurable Γ obey the pointwise bridge* (28)*, assume $\mathbb{E}|Y(a)| < \infty$ for $a \in \{0,1\}$, and let the coordinates of $\boldsymbol{\phi}_K$ be bounded. Then the projected weights in* (37) *satisfy, for every bounded measurable f,*

$$\ell_a(Z) \le w_a^\star(Y,Z) \le u_a(Z) \quad \text{on } \{D = a\}, \tag{40}$$

$$\mathbb{E}\{I_a w_a^\star(Y,Z) f(Z)\} = \mathbb{E}\{f(Z)\}, \tag{41}$$

$$\mathbb{E}\{Y(a)\} = \mathbb{E}\{I_a w_a^\star(Y,Z) Y\}. \tag{42}$$

Here $Y(a)$ denotes the potential outcome for local arm a within the eligible ordered pair. Consequently, every compatible latent historical completion produces a pair of weights in $\mathcal{W}_{1,K} \times \mathcal{W}_{0,K}$.

Every compatible historical completion therefore belongs to the observable marginal weight model. The latter also allows weight pairs whose joint latent compatibility is unrestricted, so we define it directly through the trace restrictions.

Definition C.5 (Marginal replay-drift trace model)**.** Fix the observed eligible-pair law P, the pair-level replay score q, drift radius Γ, and sieve $\boldsymbol{\phi}_K$. The marginal replay-drift model consists of the arm-mean pairs

$$\mathcal{M}_{\Gamma,K}(P,q) = \left\{ \left(\mu_1(w_1), \mu_0(w_0)\right) : \mu_a(w_a) = \mathbb{E}(I_a w_a Y), \quad w_a \in \mathcal{W}_{a,K},\ a \in \{0,1\} \right\}, \tag{43}$$

where the two arm-specific weight functions are variation independent. The corresponding causal parameter is

$$\tau_h(w_1, w_0) = \mu_1(w_1) - \mu_0(w_0), \tag{44}$$

which targets the h-higher-versus-h-lower contrast conditional on $E = 1$.

By Proposition C.4,

$$\mathcal{M}_{\mathrm{latent}}(P,q,\Gamma) \subseteq \mathcal{M}_{\Gamma,K}(P,q). \tag{45}$$

The marginal model permits more arm-mean pairs than a shared latent completion. The finite-sieve example below makes the inclusion strict. Sharpness in this section refers to Definition C.5.

Structural interpretation. The extremal arm weights are chosen separately and may depend on the observed outcome. A shared latent model additionally requires a common U, a shared $p(Z, U)$, and a joint potential-outcome law. The resulting relationship is the containment in (45).

A strict finite-sieve example. Let $Z \sim \text{Bernoulli}(1/2)$, $p = q = 1/2$, $\Gamma = \log 3$, $\phi_1 = 1$, $Y(1) = Z$, and $Y(0) = 0$. The boxes are $[4/3, 4]$. Treated-arm intercept balance gives $w(0) + w(1) = 4$, so its mean ranges over $[1/3, 2/3]$; the control mean is zero. In every compatible positive-propensity latent completion, $0 = \mathbb{P}(D = 1, Y \neq Z) = \mathbb{E}[p\mathbf{1}\{Y(1) \neq Z\}]$ forces $Y(1) = Z$, and similarly $Y(0) = 0$. Thus the latent set is $\{1/2\}$ whereas the finite-sieve marginal set is $[1/3, 2/3]$. This example establishes strict inclusion under finite-sieve balance; the full conditional-balance limit requires separate analysis.

C.4 Sharp primal characterization under the marginal trace model

Define the arm endpoints

$$\underline{\mu}_{a,K} = \inf_{w \in \mathcal{W}_{a,K}} \mathbb{E}(I_a w Y), \tag{46}$$

$$\overline{\mu}_{a,K} = \sup_{w \in \mathcal{W}_{a,K}} \mathbb{E}(I_a w Y). \tag{47}$$

The candidate causal set for the eligible-population target τ_h is

$$\mathcal{I}_{\Gamma,K} = \left[\underline{\mu}_{1,K} - \overline{\mu}_{0,K}, \overline{\mu}_{1,K} - \underline{\mu}_{0,K}\right]. \tag{48}$$

Write $\mathcal{I}_{\text{DRIFT}}(P, q; \Gamma, K) = \mathcal{I}_{\Gamma,K}$ and define the latent-model identified set

$$\mathcal{I}^{\text{lat}}_{\Gamma}(P, q) = \{\mu_1 - \mu_0 : (\mu_1, \mu_0) \in \mathcal{M}_{\text{latent}}(P, q, \Gamma)\}. \tag{49}$$

The one-way model inclusion in (45) gives

$$\mathcal{I}^{\text{lat}}_{\Gamma}(P, q) \subseteq \mathcal{I}_{\text{DRIFT}}(P, q; \Gamma, K), \qquad \tau^{\text{lat}}_h(\widetilde{P}) \in \mathcal{I}_{\text{DRIFT}}(P, q; \Gamma, K) \tag{50}$$

for every compatible latent historical completion $\widetilde{P}$.

Projection gives latent outer validity. Compactness and variation independence characterize the values attained within the marginal model.

Theorem C.6 (Sharp sieve identification under the marginal trace model)**.** *Under Assumption C.2 on the eligible-pair law, suppose the two feasible classes $\mathcal{W}_{a,K}$ are nonempty. Then they are convex and weak-$\star$ compact in $L^\infty(P)$. The infima and suprema in* (46)–(47) *are attained, and the identified set for τ_h over $\mathcal{M}_{\Gamma,K}(P, q)$ is exactly $\mathcal{I}_{\Gamma,K}$.*

Every compatible latent historical effect lies in the interval by (50). Conversely, every interval point is attained by feasible marginal weights because the two arm objective images are variation-independent intervals. This converse establishes sharpness for the marginal model, whose arm weights vary independently.

C.5 Observable dual certificates

For $r \in \mathbb{R}$, write $r_+ = \max(r, 0)$ and $r_- = \min(r, 0)$. For $\boldsymbol{\lambda} \in \mathbb{R}^K$, set

$$g^-_{a,\boldsymbol{\lambda}}(O) = \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K(Z) + I_a \left[\ell_a(Z)\{Y - \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K(Z)\}_+ + u_a(Z)\{Y - \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K(Z)\}_-\right], \tag{51}$$

$$g^+_{a,\boldsymbol{\lambda}}(O) = \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K(Z) + I_a \left[u_a(Z)\{Y - \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K(Z)\}_+ + \ell_a(Z)\{Y - \boldsymbol{\lambda}^\top \boldsymbol{\phi}_K(Z)\}_-\right]. \tag{52}$$

These scores are observable once the pair-level replay score q, the drift radius Γ, and $\boldsymbol{\phi}_K$ are specified.

Primal sharpness identifies the target set but leaves its endpoints as infinite-dimensional optimization problems. The next theorem expresses them as finite-dimensional dual values and gives inequalities that remain valid away from the optimizer.

Theorem C.7 (Strong duality and attainment)**.** *Under Assumptions C.2–C.3 on the eligible-pair law, for each arm a,*

$$\underline{\mu}_{a,K} = \sup_{\boldsymbol{\lambda}\in\mathbb{R}^K} \mathbb{E}\{g^-_{a,\boldsymbol{\lambda}}(O)\}, \tag{53}$$

$$\overline{\mu}_{a,K} = \inf_{\boldsymbol{\lambda}\in\mathbb{R}^K} \mathbb{E}\{g^+_{a,\boldsymbol{\lambda}}(O)\}. \tag{54}$$

Both extrema are attained. Moreover, for every $\boldsymbol{\lambda} \in \mathbb{R}^K$,

$$\mathbb{E}\{g^-_{a,\boldsymbol{\lambda}}(O)\} \le \underline{\mu}_{a,K} \le \overline{\mu}_{a,K} \le \mathbb{E}\{g^+_{a,\boldsymbol{\lambda}}(O)\}. \tag{55}$$

Equation (55) also supports inference with approximate optimizers. Any finite dual vector gives a directionally valid population inequality by weak duality. A vector learned on a training sample can therefore be frozen before evaluation on fresh observations; after adding a sampling correction, its score remains one-sided even if optimization is approximate. DRIFT uses the same geometry twice: Section D turns it into held-out causal certificates, while Section E differentiates the extremal value to price historical evidence.

Interpretation of the dual. The primal optimizes an outcome-dependent weight function, while the dual uses a K-dimensional vector that prices calibration constraints. Its coefficients reveal which context moments determine extremal outcomes. Derivatives of the dual value with respect to cell-specific drift radii quantify the value of historical audits in Section E.

C.6 LCE ZERO-DRIFT COROLLARY

At zero drift, exact historical provenance collapses the framework to the point-identified local contrast. The following corollary gives this nesting relation.

Corollary C.8 (LCE zero-drift recovery)**.** *Assume the clause-specific conditions of Proposition C.4 and the fixed zero-drift bridge. If $\Gamma(Z) = 0$ almost surely on the eligible-pair population, then $p_L(Z) = p_U(Z) = q(Z)$ and the feasible weights are unique on each observed arm:*

$$w_1(Y, Z) = q(Z)^{-1}, \qquad w_0(Y, Z) = \{1 - q(Z)\}^{-1}. \tag{56}$$

Consequently,

$$\mathcal{I}_{0,K} = \left\{ \mathbb{E}\left[\frac{DY}{q(Z)} - \frac{(1-D)Y}{1-q(Z)} \;\middle|\; E = 1 \right] \right\} \tag{57}$$

for every K. Thus exact ordered binary-pair LCE point identification is a strict special case of DRIFT.

At compatible zero drift, the weight envelope collapses to the LCE identity. Positive drift produces a set-valued conclusion that accounts for the discrepancy between current replay and historical selection.

C.7 INCREASING MOMENT RESOLUTION

Let $\mathcal{W}_{a,\infty}$ be the full trace-balance class on the eligible-pair law obtained by requiring

$$\mathbb{E}\{I_a w(Y, Z) f(Z)\} = \mathbb{E}\{f(Z)\} \qquad \text{for every bounded measurable } f. \tag{58}$$

Let $\underline{\mu}_{a,\infty}$ and $\overline{\mu}_{a,\infty}$ denote its endpoints.

Assumption C.9 (Sieve completeness)**.** The linear span of $\bigcup_{K\ge1}\{\boldsymbol{\phi}_K^\top b : b \in \mathbb{R}^K\}$ is dense in $L^1(P_Z)$, and the weight envelopes are uniformly bounded by a finite constant $\bar{w}$.

Zero drift collapses the envelope; under positive drift, a different route to narrower bounds is to strengthen observable balance. The next result isolates the population effect of increasing moment resolution, before the statistical section addresses its finite-sample cost.

Theorem C.10 (Nestedness and convergence of sieve certificates)**.** *Under Assumptions C.2 and C.9 on the eligible-pair law, suppose the nested classes $\mathcal{W}_{a,K}$ are nonempty for every K and both arms. Then*

$$\underline{\mu}_{a,K} \nearrow \underline{\mu}_{a,\infty}, \qquad \overline{\mu}_{a,K} \searrow \overline{\mu}_{a,\infty}, \tag{59}$$

for $a \in \{0,1\}$. Hence the intervals $\mathcal{I}_{\Gamma,K}$ are nested and converge in Hausdorff distance to

$$\mathcal{I}_{\Gamma,\infty} = [\underline{\mu}_{1,\infty} - \overline{\mu}_{0,\infty}, \overline{\mu}_{1,\infty} - \underline{\mu}_{0,\infty}]. \tag{60}$$

Increasing moment resolution removes weights inconsistent with the observed context distribution, so the population marginal interval can only contract. It remains an outer set for compatible latent historical effects. In finite samples, richer sieves can increase optimization and sampling error, making sample splitting and complexity control integral to inference.

Conditional balance and latent compatibility. Full conditional balance still permits outcome-dependent marginal weights within the bridge envelope. The resulting interval is sharp for those restrictions. Compatibility assumptions linking both potential outcomes through a shared latent state could yield a smaller structural set.

This completes the identification layer: the bridge defines admissible historical weights, calibration contracts them, and the dual certifies their extremal consequences. Held-out evaluation controls the selection bias from fitting an extremal certificate.

D Statistical Inference

Learning and evaluating a dual certificate on the same sample can make it optimistically narrow: an empirical supremum biases a lower endpoint upward, and an empirical infimum biases an upper endpoint downward. DRIFT fits each dual vector on a training subsample, freezes it, and evaluates it on held-out observations. The construction uses sample splitting as in semiparametric learning (Chernozhukov et al., 2018); its validity follows from weak duality and held-out concentration. All traces are eligible ordered pairs, and n denotes their effective sample size.

D.1 Cross-fitting and honest dual evaluation

Algorithm 1 trains dual scores and evaluates them on held-out folds using independent mechanism evidence (Figure 4). One-sided sampling corrections then form the outer confidence set.

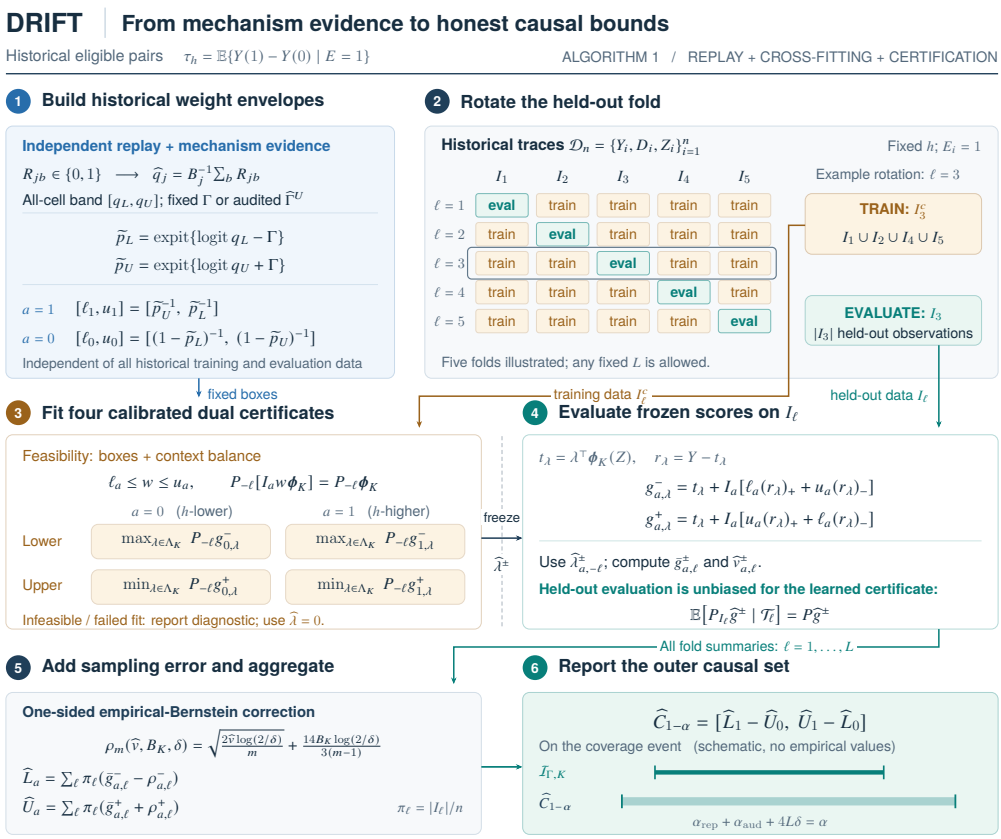

Figure 4: **DRIFT with cross-fitted dual certificates.** Replay bands and a mechanism bridge define weight boxes. Duals are trained off-fold, frozen, and evaluated with Bernstein corrections; solver failure uses the zero dual. Population-valid bands yield marginal coverage for fixed Γ, while audited radii add α_{aud} (Algorithm 1).

Figure 4 shows the training and evaluation samples used below.

Replay probabilities remain estimated quantities. In the finite-stratum design of (23),

$$\widehat{q}_j = \frac{1}{B_j} \sum_{b=1}^{B_j} R_{jb}, \qquad j = 1, \ldots, J, \tag{61}$$

and a simultaneous band $q_L(z) \le q(z) \le q_U(z)$ is constructed from these binomial samples. Because $q \mapsto \mathrm{expit}\{\mathrm{logit}(q) \pm \Gamma\}$ is increasing, a conservative historical envelope is

$$\widetilde{p}_L(z) = \mathrm{expit}\{\mathrm{logit}\, q_L(z) - \Gamma(z)\}, \qquad \widetilde{p}_U(z) = \mathrm{expit}\{\mathrm{logit}\, q_U(z) + \Gamma(z)\}. \tag{62}$$

The weight bounds are obtained from (32)–(33) after replacing (p_L, p_U) by $(\widetilde{p}_L, \widetilde{p}_U)$. On the replay-band event, these bounds contain the true drift bounds. We write $\widetilde{\mathcal{I}}_{\Gamma,K}$ for the resulting, possibly enlarged, population interval for τ_h. This enlargement accounts explicitly for replay sampling error.

For a population certificate, let the measurable band $[q_L, q_U]$ be constructed from a replay sample $\mathcal{R}$ independent of the entire historical sample, with

$$\mathbb{P}_{\mathcal{R}}\{q_L(z) \le q(z) \le q_U(z) \text{ for } P_Z\text{-almost every } z\} \ge 1 - \alpha_{\mathrm{rep}}. \tag{63}$$

Call this event $\mathcal{E}^{\mathrm{pop}}_{\mathrm{rep}}$. Independent binomial replay in every positive-mass cell of a fixed finite partition satisfies this condition when $q(z) = q_j$ throughout cell j and the cellwise bands are simultaneous. Envelopes and dual search sets must have the deterministic bounds below for every replay realization, using the known positivity restrictions when clipping. The band in (26) covers realized logged contexts; the population statement additionally requires coverage across the target law.

Proposition D.1 (Conditional replay bands do not imply population coverage)**.** *Conditional on $Z_{1:n}$, suppose each logged-context replay band has failure probability at most α_i, with $\sum_i \alpha_i \le \alpha_{\mathrm{rep}}$. Then* (26) *holds. This condition alone does not imply* (63)*.*

The outer set requires the independent population replay band. We next evaluate fitted dual scores on held-out historical observations.

Split the eligible logged traces into fixed folds $I_1, \dots, I_L$, with $m_\ell = |I_\ell|$ and $\pi_\ell = m_\ell/n$. On the complement of fold ℓ, solve

$$\widehat{\boldsymbol{\lambda}}^{-}_{a,-\ell} \in \arg\max_{\boldsymbol{\lambda}\in\Lambda_K} P_{-\ell} g^{-}_{a,\boldsymbol{\lambda}}, \qquad \widehat{\boldsymbol{\lambda}}^{+}_{a,-\ell} \in \arg\min_{\boldsymbol{\lambda}\in\Lambda_K} P_{-\ell} g^{+}_{a,\boldsymbol{\lambda}}, \tag{64}$$

where $\Lambda_K = \{\boldsymbol{\lambda} : \|\boldsymbol{\lambda}\|_2 \le R_K\}$ contains the population optimizers. Assumption C.3 and essential linear independence of the sieve imply that such a finite R_K exists; Appendix H.4 gives the coercivity argument. Evaluate the learned scores only on I_ℓ.

The fixed-population envelope bound $\bar{w}$ in Assumption C.2 and the uniform bound needed after replay inflation are distinct objects. For inference, assume that for every replay realization the inflated envelopes obey $1 \le \widetilde{\ell}_a(Z) \le \widetilde{u}_a(Z) \le \bar{w}_{\mathrm{inf}} < \infty$ for a deterministic $\bar{w}_{\mathrm{inf}}$; clipping to known replay-positivity limits is one way to enforce this condition. Under this bound and $\boldsymbol{\lambda} \in \Lambda_K$, all four replay-inflated scores are bounded by

$$\left| g^{\pm}_{a,\boldsymbol{\lambda}}(O) \right| \le B_K := R_K C_K + \bar{w}_{\mathrm{inf}}(M + R_K C_K). \tag{65}$$

Algorithm 1 DRIFT with cross-fitted dual certificates

Require: Eligible logged traces $\{Y_i, D_i, Z_i\}_{i=1}^n$ with $E_i = 1$ and candidates relabeled by h; independent replay data supporting a population-valid band; fixed deterministic drift radius Γ (or an independently audited radius with the additional error budget below); basis $\boldsymbol{\phi}_K$; fixed folds $\{I_\ell\}_{\ell=1}^L$; level α.

1: Construct $[q_L, q_U]$ from independent replay with population coverage (63), and form conservative weight envelopes using (62).
2: Check empirical feasibility of the arm-specific box and moment constraints on each training fold. If infeasible or a numerical failure prevents a fit, report the diagnostic and use the prespecified zero dual vector for that fold. Keep Γ fixed independently of this historical sample.
3: **for** $\ell = 1, \dots, L$ **do**
4: If no fallback was selected for this fold, optimize the lower and upper dual objectives in (64) on I_ℓ^c for both arms; otherwise retain the zero dual vector.
5: On I_ℓ, evaluate the four learned dual scores, their means, and their empirical variances.
6: **end for**
7: Form fold-specific one-sided empirical-Bernstein corrections and aggregate them with weights π_ℓ.
8: Return the outer causal set $[\widehat{L}_1 - \widehat{U}_0, \widehat{U}_1 - \widehat{L}_0]$ for τ_h, feasibility diagnostics, dual vectors, and cell-specific leverage scores.

For a training sigma-field $\mathcal{T}$ containing the independent replay sample, any independent pilot/audit information, and the historical training fold, and an independent eligible-pair evaluation sample of size m, write P_m^{ev} for the evaluation empirical measure.

With the replay envelope fixed, separate samples for fitting and evaluation yield the conditional expectation identities in the next proposition.

Proposition D.2 (Conditional unbiasedness and finite-sample variance)**.** *Let $O_1^{\mathrm{ev}}, \dots, O_m^{\mathrm{ev}}$ be iid from P and independent of the training sigma-field $\mathcal{T}$. Let $\widehat{\boldsymbol{\lambda}}$ be any $\mathcal{T}$-measurable dual vector and let its entire score function, including the replay envelope, be $\mathcal{T}$-measurable. The expectation on the right below integrates a fresh $O \sim P$ with the learned score held fixed. Conditional on $\mathcal{T}$,*

$$\mathbb{E}\{P_m^{\mathrm{ev}} g^{-}_{a,\widehat{\boldsymbol{\lambda}}} \mid \mathcal{T}\} = \mathbb{E} g^{-}_{a,\widehat{\boldsymbol{\lambda}}}, \tag{66}$$

$$\mathbb{E}\{P_m^{\mathrm{ev}} g^{+}_{a,\widehat{\boldsymbol{\lambda}}} \mid \mathcal{T}\} = \mathbb{E} g^{+}_{a,\widehat{\boldsymbol{\lambda}}}. \tag{67}$$

If both scores are almost surely bounded in absolute value by B_K, then

$$\mathrm{Var}\{P_m^{\mathrm{ev}} g^{\pm}_{a,\widehat{\boldsymbol{\lambda}}} \mid \mathcal{T}\} \le \frac{B_K^2}{m}. \tag{68}$$

In contrast, optimizing and evaluating the same eligible-pair sample gives

$$\mathbb{E}\sup_{\boldsymbol{\lambda}} P_m g^{-}_{a,\boldsymbol{\lambda}} \ge \sup_{\boldsymbol{\lambda}} \mathbb{E} g^{-}_{a,\boldsymbol{\lambda}}, \qquad \mathbb{E}\inf_{\boldsymbol{\lambda}} P_m g^{+}_{a,\boldsymbol{\lambda}} \le \inf_{\boldsymbol{\lambda}} \mathbb{E} g^{+}_{a,\boldsymbol{\lambda}}. \tag{69}$$

Thus the naive empirical extrema are optimistic in the direction that invalidates an outer set for τ_h.

Held-out evaluation is unbiased for the learned certificate. Its optimization gap relative to the population endpoint preserves the bound direction by weak duality. A sampling correction then turns the evaluated certificate into an outer causal statement.

D.2 Finite-sample outer confidence sets

The finite-sample result covers nonregular endpoints and approximate dual optimization. For observations $X_1, \dots, X_m \in [-B, B]$, define $\widehat{v}_m = (m-1)^{-1} \sum_{i=1}^m (X_i - \bar{X})^2$ and

$$\rho_m(\widehat{v}_m, B, \delta) = \sqrt{\frac{2\widehat{v}_m \log(2/\delta)}{m}} + \frac{14B \log(2/\delta)}{3(m-1)}. \tag{70}$$

This is the empirical-Bernstein radius for variables in $[-B, B]$ (Maurer & Pontil, 2009). Let $\bar{g}^{\pm}_{a,\ell}$ and $\widehat{v}^{\pm}_{a,\ell}$ be the evaluation mean and variance in fold ℓ. Define

$$\widehat{L}_a = \sum_{\ell=1}^{L} \pi_\ell \left\{ \bar{g}^{-}_{a,\ell} - \rho_{m_\ell}(\widehat{v}^{-}_{a,\ell}, B_K, \delta) \right\}, \tag{71}$$

$$\widehat{U}_a = \sum_{\ell=1}^{L} \pi_\ell \left\{ \bar{g}^{+}_{a,\ell} + \rho_{m_\ell}(\widehat{v}^{+}_{a,\ell}, B_K, \delta) \right\}. \tag{72}$$

Applying empirical-Bernstein corrections to the held-out scores and combining the four armwise directions with the population replay event yields the primary inferential guarantee.

Theorem D.3 (Finite-sample valid outer causal set)**.** *Let the historical observations be iid from the eligible ordered-pair law P, with folds fixed independently of those observations. Fix a deterministic radius Γ. Assume bounded outcomes and basis as in Assumption C.2, and suppose the fixed-population classes $\mathcal{W}_{a,K}$ induced by (P, q, Γ) are nonempty. Let the replay band be independent of the entire historical sample and satisfy* (63)*. Assume $m_\ell \ge 2$, and assume the replay-inflated envelopes and every returned dual vector obey the deterministic uniform bounds $\bar{w}_{\inf}$ and R_K, hence the score bound B_K in* (65)*, for every replay realization. Set*

$$\delta = \frac{\alpha - \alpha_{\mathrm{rep}}}{4L}, \qquad 0 \le \alpha_{\mathrm{rep}} < \alpha. \tag{73}$$

Then, for arbitrary measurable training algorithms used to select the fold-specific dual vectors,

$$\mathbb{P}\left\{ \mathcal{I}_{\Gamma,K}(P, q) \subseteq \widetilde{\mathcal{I}}_{\Gamma,K} \subseteq [\widehat{L}_1 - \widehat{U}_0,\ \widehat{U}_1 - \widehat{L}_0] \right\} \ge 1 - \alpha. \tag{74}$$

The guarantee does not require the dual programs to be solved exactly: every returned dual vector in the bounded search set gives a valid, potentially looser, certificate for τ_h. Moreover, if the conditions of Proposition C.4 hold with this fixed Γ, then for every compatible latent historical completion $\widetilde{P}$,

$$\mathbb{P}\left\{ \tau_h^{\mathrm{lat}}(\widetilde{P}) \in [\widehat{L}_1 - \widehat{U}_0,\ \widehat{U}_1 - \widehat{L}_0] \right\} \ge 1 - \alpha. \tag{75}$$

Corollary D.4 (Finite-sample certification at the zero-drift boundary)**.** *Let the historical observations be iid from the eligible ordered-pair law, and fix the folds independently of those observations. For each fold, require every dual vector and its entire score function to be trained on the fold complement and evaluated only on the held-out fold. Assume consistency and historical conditional randomization, let the known fixed replay score satisfy $p(Z, U) = q(Z) \in [\eta_q, 1 - \eta_q]$ almost surely,*

and assume $0 \le Y \le M$ and $\|\phi_K(Z)\|_2 \le C_K$. Use folds with $m_\ell \ge 2$, the collapsed envelopes at $\Gamma = 0$, and any training rule whose returned dual scores satisfy a deterministic bound $B_{K,0}$. With $\delta = \alpha/(4L)$, form (71)–(72) using $B_{K,0}$. Then

$$\mathbb{P}\Big\{\mathcal{I}_{0,K} = \{\tau_h\} \subseteq [\widehat{L}_1 - \widehat{U}_0, \widehat{U}_1 - \widehat{L}_0]\Big\} \ge 1 - \alpha.$$

This boundary result uses feasibility of the collapsed boxes, weak duality, and held-out concentration; it does not invoke Assumption C.3.

Corollary D.5 (Finite-sample certification with an audited random radius)**.** *Let the historical observations be iid from the eligible ordered-pair law P, fix the folds independently of those observations, and train every fold-specific dual and score function only on the fold complement before evaluating it on the held-out fold. Assume $0 \le Y \le M$ and $\|\phi_K(Z)\|_2 \le C_K$. Let $\Gamma^\star$ be a fixed measurable drift radius satisfying the pointwise bridge for the compatible latent completions under consideration. Let $\widehat{\Gamma}^U$ be a random upper radius constructed from pilot/audit information, and suppose $\mathbb{P}(\mathcal{E}_{\mathrm{aud}}) \ge 1 - \alpha_{\mathrm{aud}}$ for the event in (31). Let a population-valid replay band satisfy (63). The sigma-field generated jointly by all pilot, audit, and replay information must be independent of the entire historical sample used for dual training and evaluation; no mutual independence between the audit event and replay-band event is required. Assume the random replay-and-audit-inflated envelopes and all returned dual vectors obey deterministic uniform bounds $\bar{w}_{\mathrm{inf}}^{\mathrm{aud}}$ and R_K for every realization, and define the corresponding all-realization score bound*

$$B_K^{\mathrm{aud}} = R_K C_K + \bar{w}_{\mathrm{inf}}^{\mathrm{aud}}(M + R_K C_K). \tag{76}$$

Suppose the fixed population classes induced by $(P, q, \Gamma^\star)$ are nonempty and $m_\ell \ge 2$. Set

$$\delta = \frac{\alpha - \alpha_{\mathrm{rep}} - \alpha_{\mathrm{aud}}}{4L} > 0. \tag{77}$$

Then the certificates constructed with $\widehat{\Gamma}^U$ and B_K^{aud} obey

$$\mathbb{P}\Big\{\mathcal{I}_{\Gamma^\star,K}(P,q) \subseteq \widetilde{\mathcal{I}}_{\widehat{\Gamma}^U,K} \subseteq [\widehat{L}_1 - \widehat{U}_0, \widehat{U}_1 - \widehat{L}_0]\Big\} \ge 1 - \alpha.$$

If the conditions of Proposition C.4 hold for $\Gamma^\star$, the same event covers every compatible latent historical effect. A radius tuned on any part of the historical training or evaluation sample is outside this guarantee, even when its tuning rule was prespecified.

Population coverage and held-out evaluation. The theorem and its two corollaries cover causal sets. The replay band protects population weight-class inclusion, while held-out Bernstein corrections protect learned dual expectations. Approximate optimization can enlarge a certificate while preserving its one-sided direction. Positive-mass contexts absent from realized history still contribute to the target, so replay bands must cover the population rather than only observed contexts. Conditional on independent replay or combined replay/audit information and a fold's training observations, evaluation traces remain iid from P. Integrating foldwise bounds and applying a union bound gives joint coverage even though the folds may depend on one another.

Optimization accuracy. Every bounded dual vector gives a one-sided population bound, so approximate optimization can widen the certificate while preserving coverage. Under the stable unique-optimizer conditions of Theorem D.7, optimizer estimation contributes only a second-order term.

D.3 Joint historical/replay central limit theorem

Under additional endpoint stability, the cross-fitted scores admit a two-sample expansion that separates historical-outcome uncertainty from uncertainty in the replay anchor.

Let $\boldsymbol{\lambda}_a^{-,\star}$ and $\boldsymbol{\lambda}_a^{+,\star}$ be the unique population optimizers in (53)–(54) for the eligible-pair law. Define

$$S_K(O) = \Big(g^-_{1,\boldsymbol{\lambda}_1^{-,\star}}, g^+_{1,\boldsymbol{\lambda}_1^{+,\star}}, g^-_{0,\boldsymbol{\lambda}_0^{-,\star}}, g^+_{0,\boldsymbol{\lambda}_0^{+,\star}}\Big)^\top, \qquad \Sigma_K = \mathrm{Var}\{S_K(O)\}. \tag{78}$$

In this subsection, $q = (q_1, \ldots, q_J)$ denotes the finite replay-probability vector. Index the four arm endpoints by $e = (a, \pm)$ and write g_e for the corresponding dual score.

Assumption D.6 (Locally stable two-sample dual optimizers)**.** For fixed K and J, there is a neighborhood $\mathcal{Q}$ of the true replay vector q, contained in the positivity interior, such that each objective $G_e(\boldsymbol{\lambda}, q') = Pg_e(O; \boldsymbol{\lambda}, q')$ has a unique optimizer $\boldsymbol{\lambda}_e^\star(q')$ in a common compact subset of the interior of Λ_K. Each G_e is jointly twice continuously differentiable near these optimizers; its Hessian in $\boldsymbol{\lambda}$ is uniformly negative definite for a lower endpoint and uniformly positive definite for an upper endpoint. The envelopes, scores, and their first derivatives in q' have common deterministic bounds. There is no probability mass at $Y = \boldsymbol{\lambda}_e^\star(q')^\top \boldsymbol{\phi}_K(Z)$ on arm a for $q' \in \mathcal{Q}$. Training at the replay estimate satisfies, for each fold,

$$\|\widehat{\boldsymbol{\lambda}}_{e,-\ell}(\widehat{q}) - \boldsymbol{\lambda}_e^\star(\widehat{q})\|_2 = O_p(n^{-1/2}).$$

This last requirement is local in the estimated replay argument; fixed-q consistency alone is insufficient. Here and in the CLT, the raw replay means are clipped to $[\eta_q/2, 1 - \eta_q/2]$, where the true probabilities lie in $[\eta_q, 1 - \eta_q]$; clipping is asymptotically inactive.

Residual mass near the fitted threshold provides a diagnostic for the no-kink-mass condition, which supports local regularity. When this condition fails, endpoint limits may be non-Gaussian; Theorem D.3 still supplies an outer certificate. Asymptotic normality gives a refinement in stable regimes.

Let

$$\widehat{\mu}_{a,K}^{\pm} = \sum_{\ell=1}^{L} \pi_\ell P_{I_\ell} g_{a,\widehat{\boldsymbol{\lambda}}_{a,-\ell}^{\pm}}^{\pm} \tag{79}$$

be the cross-fitted score averages without Bernstein expansion. In the two-sample formulation, these scores use the replay-estimated envelope built from $\widehat{q}$ in (61); we write the result as $\widehat{\mu}_{a,K}^{\pm}(\widehat{q})$ when the replay argument needs to be displayed, and write $\mu_{a,K}^{\pm}(q)$ for its population counterpart.

Theorem D.7 (Joint two-sample asymptotic normality of marginal-model sharp endpoints)**.** *Under Assumptions C.2, C.3, and D.6 on the eligible-pair law, fixed K, a fixed number of asymptotically balanced folds, the independent sampling design* (23)*, and a Hadamard-differentiable endpoint map in the finite vector $(q_1, \ldots, q_J)$, define*

$$\widehat{\theta} = \left(\widehat{\mu}_{1,K}^{-}(\widehat{q}), \widehat{\mu}_{1,K}^{+}(\widehat{q}), \widehat{\mu}_{0,K}^{-}(\widehat{q}), \widehat{\mu}_{0,K}^{+}(\widehat{q})\right)^\top, \qquad \theta = \left(\underline{\mu}_{1,K}, \overline{\mu}_{1,K}, \underline{\mu}_{0,K}, \overline{\mu}_{0,K}\right)^\top. \tag{80}$$

There are centered historical and replay influence functions ψ_O and $\psi_{R,j}$ such that

$$\widehat{\theta} - \theta = \frac{1}{n}\sum_{i=1}^{n} \psi_O(O_i) + \frac{1}{B}\sum_{j=1}^{J}\sum_{b=1}^{B_j} \psi_{R,j}(R_{jb}) + o_p\left(n^{-1/2} + B^{-1/2}\right). \tag{81}$$

Equivalently, after flattening the replay draws, the second term is $B^{-1}\sum_{b=1}^{B} \psi_R(R_b)$. If $\dot{\theta}_{q,j}$ denotes the derivative with respect to q_j, one may take

$$\psi_{R,j}(R_{jb}) = \rho_j^{-1}\dot{\theta}_{q,j}\{R_{jb} - q_j\}, \qquad \rho_j = \lim B_j/B. \tag{82}$$

Writing $V_O = \mathrm{Var}\{\psi_O(O)\}$ and $V_R = \sum_{j=1}^{J} \rho_j \,\mathrm{Var}\{\psi_{R,j}(R_j)\}$, independence of $\mathcal{D}_n$ and $\mathcal{R}_B$ gives

$$\mathrm{Avar}(\widehat{\theta}) := \frac{V_O}{n} + \frac{V_R}{B}. \tag{83}$$

Therefore, if $B/n \to \kappa \in (0, \infty)$, $\sqrt{n}(\widehat{\theta} - \theta)$ converges to $N(0, V_O + \kappa^{-1}V_R)$, and the known-q result is recovered by setting $V_R = 0$. The empirical influence contributions consistently estimate the displayed leading asymptotic covariance after the corresponding sample-size scaling. This is an asymptotic covariance statement; convergence of the estimator's actual second moments would additionally require an L^2 remainder or uniform-integrability condition. Writing $S_{a,\pm}$ for the corresponding coordinate of S_K, the historical score contrasts for the lower and upper endpoints of τ_h are

$$S_{\tau_h,L} = S_{1,-} - S_{0,+}, \qquad S_{\tau_h,U} = S_{1,+} - S_{0,-}. \tag{84}$$

Centering these contrasts gives their historical influence functions. The replay components are the same linear combinations of the corresponding $\psi_{R,j}$ terms in (81)*.*

For $r_e = Y - \boldsymbol{\lambda}_e^\star(q)^\top \boldsymbol{\phi}_K(Z)$ and cell S_j, the replay derivatives are (using the corresponding lower or upper residual in each line)

$$d_{a,j}^- = \mathbb{E}[I_a \mathbf{1}\{Z \in S_j\}\{\ell_a'(q_j)(r_e)_+ + u_a'(q_j)(r_e)_-\}], \tag{85}$$

$$d_{a,j}^+ = \mathbb{E}[I_a \mathbf{1}\{Z \in S_j\}\{u_a'(q_j)(r_e)_+ + \ell_a'(q_j)(r_e)_-\}], \tag{86}$$

where derivatives are pointwise in Z when Γ varies within a cell:

$$\ell_1' = -e^{-\Gamma}/q^2, \quad u_1' = -e^{\Gamma}/q^2, \qquad \ell_0' = e^{-\Gamma}/(1-q)^2, \quad u_0' = e^{\Gamma}/(1-q)^2.$$

Let d_j stack the four derivatives. Cross-fitted averages of these derivative integrands, with the fitted dual and replay argument substituted, consistently estimate d_j. The replay covariance estimator is

$$\widehat{V}_R = \sum_j \frac{\widehat{d}_j \widehat{d}_j^\top \widehat{q}_j(1-\widehat{q}_j)}{B_j/B}.$$

The empirical covariance of the cross-fitted score vectors estimates V_O. Confidence-band inflation is not used in this plug-in CLT estimator.

First-order optimality and cross-fitting make optimizer estimation second order under fixed K, fixed J, and the stated stability conditions. The expansion contains separate historical and replay influence contributions for the specified two-sample experiment. Efficient sharp endpoint estimators for other sensitivity and robust OPE models provide related comparisons (Bennett et al., 2024; Hess et al., 2026; Zhang & Zhao, 2026).

The Gaussian expansion requires active-set stability in addition to positivity and primal feasibility. The following construction shows what changes when an endpoint optimizer lies at a kink.

Proposition D.8 (A nonregular active-set boundary)**.** *Strict primal feasibility and historical positivity do not guarantee Gaussian endpoint estimation. Consider one replay cell, intercept-only balance, known $q = 1/2$, $\Gamma = \log 2$, historical $D \sim \mathrm{Bernoulli}(3/5)$, $Y(0) = 0$, and $Y(1) \sim \mathrm{Bernoulli}(8/9)$ independently of D. The marginal treated lower endpoint is $4/5$. Its empirical optimized value $\widehat{\mu}_1^-$, defined arbitrarily on empirically infeasible samples, satisfies*

$$\sqrt{n}(\widehat{\mu}_1^- - 4/5) \;\Rightarrow\; A + |H|, \qquad A \sim N(0, 9/25), \quad H \sim N(0, 1/5), \quad A \perp H.$$

This limit is non-Gaussian. The example violates the stable-optimizer assumptions of Theorem D.7; the finite-sample certificate of Theorem D.3 still applies.

The non-Gaussian component is first order at this boundary. The outer certificate remains available, while Gaussian endpoint rates apply under the explicit regularity conditions above.

D.4 Sampling rates and identification ambiguity

We study two risks: estimation of the marginal identified interval from historical and replay samples, and point decisions within that interval. Point decisions can remain ambiguous with a known observable law. This ambiguity is defined over the marginal model, whose endpoints need not both admit compatible latent historical completions.

Let d_H denote Hausdorff distance between compact intervals. The following explicit, positive-drift model supplies both lower and upper expected squared-risk bounds on the same parameter class. Its outcomes take values in $\{0, M\}$; replay uses the independent one-stratum Bernoulli design.

Theorem D.9 (Matching sampling rates and marginal decision ambiguity)**.** *Fix $M > 0$ and $0 < \eta < 1/2$. Choose $\epsilon > 0$ sufficiently small that $[1/2 - \epsilon, 1/2 + \epsilon] \subset [\eta, 1-\eta]$ and $\epsilon \le 1/100$. Consider the one-cell, intercept-only marginal model with certified radius $\Gamma_\star = \log 2$, historical propensity $p = 1/2$, $Y(0) = 0$, $Y(1)/M \sim \mathrm{Bernoulli}(t)$ independently of display, and current replay probability q, where $(t, q) \in \mathcal{C}_\epsilon = [1/2 - \epsilon, 1/2 + \epsilon]^2$. Write $\ell(q) = (1+q)/(2q)$. The marginal identified interval is*

$$\mathcal{I}(t, q) = \left[\frac{Mt\ell(q)}{2}, M\left\{1 - \frac{(1-t)\ell(q)}{2}\right\}\right], \qquad r(q) = \frac{M}{2}\left\{1 - \frac{\ell(q)}{2}\right\} > 0.$$

For n independent historical observations and B independent replay draws, there are constants $0 < c < C < \infty$ depending only on (M, ϵ) such that

$$c(n^{-1} \vee B^{-1}) \le \inf_{\widehat{\mathcal{I}}} \sup_{\mathcal{C}_\epsilon} E d_H^2(\widehat{\mathcal{I}}, \mathcal{I}(t,q)) \le C(n^{-1} + B^{-1}).$$

Let $r_{\min} = \min_q r(q) > 0$ and $r_{\max} = \max_q r(q)$. For the marginal point-decision problem on this same class,

$$c' \max\{n^{-1}, B^{-1}, r_{\min}^2\} \le \inf_{T} \sup_{\mathcal{C}_\epsilon} \sup_{\theta \in \mathcal{I}(t,q)} E(T-\theta)^2 \le C'(n^{-1} + B^{-1} + r_{\max}^2).$$

At any fixed observable law with interval $[L, U]$ and radius $r = (U-L)/2$, the additional exact decision bound is

$$\inf_{T} \sup_{\theta \in [L,U]} \mathbb{E}(T-\theta)^2 \ge r^2. \tag{87}$$

Here $r_{\max}/r_{\min}$ is bounded on the specified class. The sampling rates match uniformly on this compact finite-atom class.

Meaning of the lower bound. The $n^{-1/2}$ scale reflects outcome sampling among eligible pairs. The $B^{-1/2}$ scale reflects uncertainty in the replay anchor. The radius $r_{\min}$ measures marginal-model decision ambiguity. Because $\mathcal{I}_\Gamma^{\mathrm{lat}}(P,q) \subseteq \mathcal{I}_{\Gamma,K}(P,q)$ may be strict, an $r_{\min}^2$ lower bound for compatible latent historical completions would require additional structural compatibility.

The projected plug-in estimator in the proof has uniformly bounded expected squared endpoint error on the explicit compact class, establishing the matching upper rate. This expected-risk result uses the compact-class argument in the proof. The ambiguity term motivates mechanism audits that can reduce the certified radius.

E Mechanism Audits and First-Order Allocation

Historical mechanism evidence can tighten a sensitivity analysis. Restoring an archived selector, replaying an old retrieval snapshot, collecting overlap-period shadow traffic, or reconstructing a filter state may constrain historical selection in particular contexts. These audits differ in cost and inferential value. We study how to allocate them across eligible ordered-pair contexts to reduce the width of the marginal identified set $\mathcal{I}_{\mathbf{\Gamma},K}$, which remains an outer bound for compatible latent effects.

We allocate audits to reduce marginal identification width. Current replay improves precision about q, and historical audits constrain the bridge to p. Three quantities determine priority: s_j measures how strongly cell j changes the extremal causal conclusion, c_j measures how quickly evidence reduces its bridge uncertainty, and κ_j records acquisition cost.

E.1 Audit value from the dual derivative

Let

$$W(\mathbf{\Gamma}) = \mathrm{length}\{\mathcal{I}_{\mathbf{\Gamma},K}\} \tag{88}$$

be the population causal width for the eligible-population target τ_h. At a reference radius $\mathbf{\Gamma}^0$, define the one-sided dual causal leverage

$$s_j = \left.\frac{\partial W(\mathbf{\Gamma})}{\partial \Gamma_j}\right|_{\mathbf{\Gamma}=\mathbf{\Gamma}^0} \ge 0. \tag{89}$$

The first design question is which cells can change the causal interval at all. The following proposition answers it locally by differentiating the exact marginal width through its four dual endpoint representations.

Proposition E.1 (Derivative of the exact marginal width)**.** *Fix a finite cell partition $C_1, \ldots, C_J$, P, q, K, bounded outcomes and basis, and a reference vector $\mathbf{\Gamma}^0$. Suppose q is bounded away from zero and one, all radii are bounded in a relative neighborhood of $\mathbf{\Gamma}^0$, the four dual representations hold there, and their extrema are attained in one common compact set of dual vectors. Suppose the*

four dual optimizers at $\mathbf{\Gamma}^0$ are unique. Write $r_a^{\pm} = Y - (\boldsymbol{\lambda}_a^{\pm})^\top \boldsymbol{\phi}_K(Z)$ for their residuals. Then the exact marginal width is differentiable at $\mathbf{\Gamma}^0$ (relative to nonnegative admissible directions at a boundary), and

$$s_j = \sum_{a=0}^{1} \mathbb{E}\Big[I_a \mathbf{1}\{Z \in C_j\} \big\{ (u_a - 1)(r_a^+)_+ - (\ell_a - 1)(r_a^+)_- \tag{90}$$

$$+ (\ell_a - 1)(r_a^-)_+ - (u_a - 1)(r_a^-)_- \big\} \Big] \geq 0. \tag{91}$$

All weights and optimizers in the formula are evaluated at $\mathbf{\Gamma}^0$. Uniform strict feasibility with bounded boxes and nondegenerate moments is one sufficient route to the common compact optimizer set. Uniqueness alone is not a substitute for this local attainment/stability condition.

At nonregular points, one-sided numerical differences give exploratory allocation scores. A supergradient certificate requires an additional bound.

For the tutoring application, an audit has high value when recovering its historical selector changes an extremal conclusion about retention. Its derivative depends jointly on outcomes and calibration, through the same dual programs used for inference. The next step relates this value to the rate at which audits constrain the bridge.

E.2 An auditable cell-level bridge radius

We first make the bridge radius an estimable object. In cell j, let an archived historical-mechanism audit and a current replay experiment record independent binary draws

$$H_{jk} \sim \text{Bernoulli}(p_j), \quad k = 1, \ldots, m_j, \qquad R_{jb} \sim \text{Bernoulli}(q_j), \quad b = 1, \ldots, B_j, \tag{92}$$

where one indicates selection of the h-higher candidate from the same eligible ordered pair. To turn a cell probability into the pointwise bridge required by Assumption C.1, impose cell homogeneity: $p(Z, U) = p_j$ and $q(Z) = q_j$ almost surely throughout cell j. Heterogeneous cells require a separately certified bound on within-cell log-odds variation to convert cell-average audit probabilities into a pointwise bridge. Assume $p_j, q_j \in [\eta, 1 - \eta]$ for a fixed $\eta > 0$. Write $\widehat{p}_j = m_j^{-1} \sum_k H_{jk}$ and $\widehat{q}_j = B_j^{-1} \sum_b R_{jb}$. For a simultaneous audit level α_{aud}, define

$$e_{p,j} = \sqrt{\frac{\log(4J/\alpha_{\text{aud}})}{2m_j}}, \qquad e_{q,j} = \sqrt{\frac{\log(4J/\alpha_{\text{aud}})}{2B_j}}, \tag{93}$$

and, with $\eta_0 = \eta/2$ and $\text{clip}_{\eta_0}(x) = \min\{1 - \eta_0, \max\{\eta_0, x\}\}$,

$$\begin{aligned} p_j^L &= \text{clip}_{\eta_0}(\widehat{p}_j - e_{p,j}), & p_j^U &= \text{clip}_{\eta_0}(\widehat{p}_j + e_{p,j}), \\ q_j^L &= \text{clip}_{\eta_0}(\widehat{q}_j - e_{q,j}), & q_j^U &= \text{clip}_{\eta_0}(\widehat{q}_j + e_{q,j}). \end{aligned} \tag{94}$$

The clipping only keeps the logit finite; it preserves coverage because the true probabilities lie in $[\eta, 1 - \eta]$. Define the auditable upper bridge radius

$$\widehat{\Gamma}_j^U = \max_{p \in [p_j^L, p_j^U],\, q \in [q_j^L, q_j^U]} |\text{logit}\, p - \text{logit}\, q| = \max\left\{ \left|\text{logit}\, p_j^L - \text{logit}\, q_j^U\right|, \left|\text{logit}\, p_j^U - \text{logit}\, q_j^L\right| \right\}. \tag{95}$$

The Bernoulli experiments above quantify how quickly evidence shrinks a cell's bridge radius. The next theorem converts their simultaneous probability bands into a finite-sample bridge certificate.

Theorem E.2 (Finite-sample audit certificate for the bridge radius)**.** *Fix the cells before outcome analysis and suppose $p(Z, U) = p_j$ and $q(Z) = q_j$ almost surely within cell j, with $p_j, q_j \in [\eta, 1 - \eta]$. Conditional on an independent pilot sigma-field, let the final counts m_j, B_j be fixed, and collect fresh historical-audit and current-replay draws satisfying* (92)*. Conditional on the pilot, assume the final Bernoulli innovations are independent within and across cells, independent between the historical-audit and replay experiments, and fresh relative to the pilot. For subsequent use in Corollary D.5, additionally require the pilot and all final audit/replay draws to*

be jointly independent of the entire reserved historical inference sample. Then, conditionally on the pilot and hence also unconditionally, with probability at least $1-\alpha_{\mathrm{aud}}$ *simultaneously for all* j*,*

$$p_j \in [p_j^L, p_j^U], \qquad q_j \in [q_j^L, q_j^U], \qquad \Gamma_j := |\mathrm{logit}\, p_j - \mathrm{logit}\, q_j| \le \widehat{\Gamma}_j^U. \tag{96}$$

Moreover, for $L_{\eta_0} = 1/\{\eta_0(1-\eta_0)\}$*, on the same simultaneous coverage event,*

$$0 \le \widehat{\Gamma}_j^U - \Gamma_j \le 2L_{\eta_0}(e_{p,j} + e_{q,j}) = O\Big(\sqrt{\log(4J/\alpha_{\mathrm{aud}})}\,[m_j^{-1/2} + B_j^{-1/2}]\Big). \tag{97}$$

If, in addition, $\alpha_{\mathrm{aud}} = \alpha_{\mathrm{aud},m,C_{\mathrm{aud}}} \to 0$ *and*

$$\log(J/\alpha_{\mathrm{aud},m,C_{\mathrm{aud}}}) = o\Big(\min_j\{m_j, B_j\}\Big), \tag{98}$$

then the absolute estimation error satisfies

$$\Big|\widehat{\Gamma}_j^U - \Gamma_j\Big| = O_p\Big(\sqrt{\log(4J/\alpha_{\mathrm{aud}})}\,[m_j^{-1/2} + B_j^{-1/2}]\Big). \tag{99}$$

For fixed J *and fixed* α_{aud}*, the absolute error is* $O_p(m_j^{-1/2} + B_j^{-1/2})$ *directly from concentration of the two sample means and the deterministic band widths.*

The two observable Bernoulli experiments thus certify the bridge radius. For fixed J and fixed α_{aud}, if the replay-to-audit ratio is fixed, $B_j = \beta_j m_j$ with $\beta_j \in (0,\infty)$, define the deterministic comparison radius used in the design analysis by

$$\Gamma_j^{\mathrm{cert}}(m_j) = \Gamma_{j,\infty} + c_j m_j^{-1/2}, \qquad \Gamma_{j,\infty} = \Gamma_j, \qquad c_j = 2L_{\eta_0}\sqrt{\frac{\log(4J/\alpha_{\mathrm{aud}})}{2}}\left(1+\beta_j^{-1/2}\right). \tag{100}$$

On the event in (96), $\widehat{\Gamma}_j^U \le \Gamma_j^{\mathrm{cert}}(m_j)$. The comparison curve depends on the unknown irreducible drift Γ_j; the implemented certificate is $\widehat{\Gamma}_j^U$. The one-term allocation rate below uses fixed confidence and proportional replay budgets. If the confidence level varies, its logarithmic factor remains part of the rate.

E.3 First-order optimal allocation

Partition the logged eligible-pair context into J bridge cells. Let m_j be the number of historical audits in cell j and $\kappa_j > 0$ the cost per audit. Under the auditable experiment above with a fixed replay-to-audit ratio, analyze the deterministic comparison curve

$$\Gamma_j^{\mathrm{cmp}}(m_j) = \Gamma_{j,\infty} + c_j m_j^{-1/2}, \qquad c_j > 0, \tag{101}$$

where $\Gamma_{j,\infty}$ is the irreducible historical-versus-current log-odds discrepancy and c_j is the explicit audit/replay uncertainty scale in (100). The square-root comparison scale follows from the Bernoulli certificate bound above.

The allocation results optimize the deterministic comparison width $W\{\mathbf{\Gamma}^{\mathrm{cmp}}(m)\}$. The realized random certificate width $W(\widehat{\mathbf{\Gamma}}^U)$ has a separate distribution. Set $s_j = \partial_j W(\mathbf{\Gamma}_\infty)$ at the irreducible radius. The unit cost κ_j includes one historical audit and its β_j replay draws when both are budgeted; alternatively replay is funded separately and C_{aud} is the historical-audit budget. Independent pilot cost is excluded from the remaining budget C_{aud}; a total-budget asymptotic claim requires pilot cost $o(C_{\mathrm{aud,total}})$.

The first-order width approximation is

$$\overline{W}(m_1,\ldots,m_J) = W(\mathbf{\Gamma}_\infty) + \sum_{j=1}^{J}\frac{a_j}{\sqrt{m_j}}, \qquad a_j = s_j c_j. \tag{102}$$

Equation (102) is a first-order comparison objective; its Taylor remainder has no fixed sign. The product a_j is the design coefficient: c_j measures bridge uncertainty, while s_j measures how resolving it changes the causal conclusion for τ_h.

Algorithm 2 Mechanism audits using dual derivatives

Require: A pilot/audit design independent of the historical sample reserved for Algorithm 1; a consistent pilot estimate of $\mathbf{\Gamma}_\infty$, pilot estimates $\widehat{s}_j, \widehat{c}_j$, costs κ_j, budget C_{aud}, minimum cell audits $m_{j,0}$, replay counts B_j, audit level α_{aud}, and eligible ordered-pair traces.

1: Fit the DRIFT dual programs on independent pilot traces at the consistent estimated irreducible radius. Consistent derivatives at this reference are required by Theorem E.5; derivatives at a fixed distant radius are only heuristic.
2: **for** $j = 1, \ldots, J$ **do**
3: Compute $\widehat{s}_j$ by an analytic dual derivative or a one-sided finite difference that re-solves only the affected certificate, and compute $\widehat{c}_j$ from pilot audit/replay data. For numerical derivatives, the step size and pilot error must ensure derivative consistency; design estimates are not conservative certificates by themselves.
4: **end for**
5: Set $\widehat{a}_j = (\widehat{s}_j)_+(\widehat{c}_j)_+$ and allocate the residual budget by (110), enforcing lower bounds through the KKT active set.
6: Conditional on the pilot, fix every final count, round down within budget, and collect fresh independent H_{jk} and R_{jb} draws for the same eligible ordered pairs. Complete all final counts, then form (94), compute $\widehat{\Gamma}_j^U$ by (95), and run Algorithm 1 on the reserved historical sample with these radii and the total failure budget in (77).
7: Return the completed fixed-sample certificates. Repeated inspection with outcome-dependent stopping requires a separate sequential coverage guarantee.

Theorem E.3 (Closed-form optimal bridge allocation)**.** *Let $a_j > 0$, $\kappa_j > 0$, and $C_{\mathrm{aud}} > 0$, where C_{aud} in this section denotes total audit cost budget, while B_j denotes cell replay counts. The unique solution to*

$$\min_{m_j > 0} \sum_{j=1}^{J} a_j m_j^{-1/2} \quad \textit{subject to} \quad \sum_{j=1}^{J} \kappa_j m_j \le C_{\mathrm{aud}} \tag{103}$$

uses the full budget and is

$$m_j^\star = \frac{C_{\mathrm{aud}}\,(a_j/\kappa_j)^{2/3}}{\sum_{k=1}^{J} a_k^{2/3}\kappa_k^{1/3}} = \frac{C_{\mathrm{aud}}\,(s_j c_j/\kappa_j)^{2/3}}{\sum_{k=1}^{J} (s_k c_k)^{2/3}\kappa_k^{1/3}}. \tag{104}$$

The minimum first-order excess width is

$$\min_{m} \sum_{j} a_j m_j^{-1/2} = \frac{\left(\sum_{j=1}^{J} a_j^{2/3}\kappa_j^{1/3}\right)^{3/2}}{\sqrt{C_{\mathrm{aud}}}}. \tag{105}$$

With lower bounds $m_j \ge m_{j,0}$, the same formula applies to the residual budget after the Karush–Kuhn–Tucker active set is determined.

The closed form optimizes the deterministic first-order width. The next two results state when it also approximates the exact width W and can be estimated from a pilot. Figure 5 summarizes Algorithm 2: independent pilot estimates of causal leverage and audit uncertainty determine a cost-constrained allocation, and fresh fixed-count experiments certify the bridge before inference on reserved historical data. The worked example distinguishes the number of audits from their cost.

Figure 5 shows the independence structure used by Algorithm 2.

Corollary E.4 (First-order optimality for the exact causal width)**.** *Assume J is fixed, W is Fréchet differentiable at $\mathbf{\Gamma}_\infty$ from the positive orthant, and*

$$W(\mathbf{\Gamma}_\infty + \boldsymbol{h}) = W(\mathbf{\Gamma}_\infty) + \sum_{j=1}^{J} s_j h_j + o(\|\boldsymbol{h}\|_2) \tag{106}$$

uniformly over nonnegative directions, where $s_j = \partial_j W(\mathbf{\Gamma}_\infty)$ and $s_j c_j > 0$ for every j. Let $\Gamma_j^{\mathrm{cmp}}(m_j) = \Gamma_{j,\infty} + c_j m_j^{-1/2}$ and permit continuous audit allocations. As $C_{\mathrm{aud}} \to \infty$, the allocation in (104)*, with $a_j = s_j c_j$, is asymptotically optimal for the* exact marginal-model width *of the*

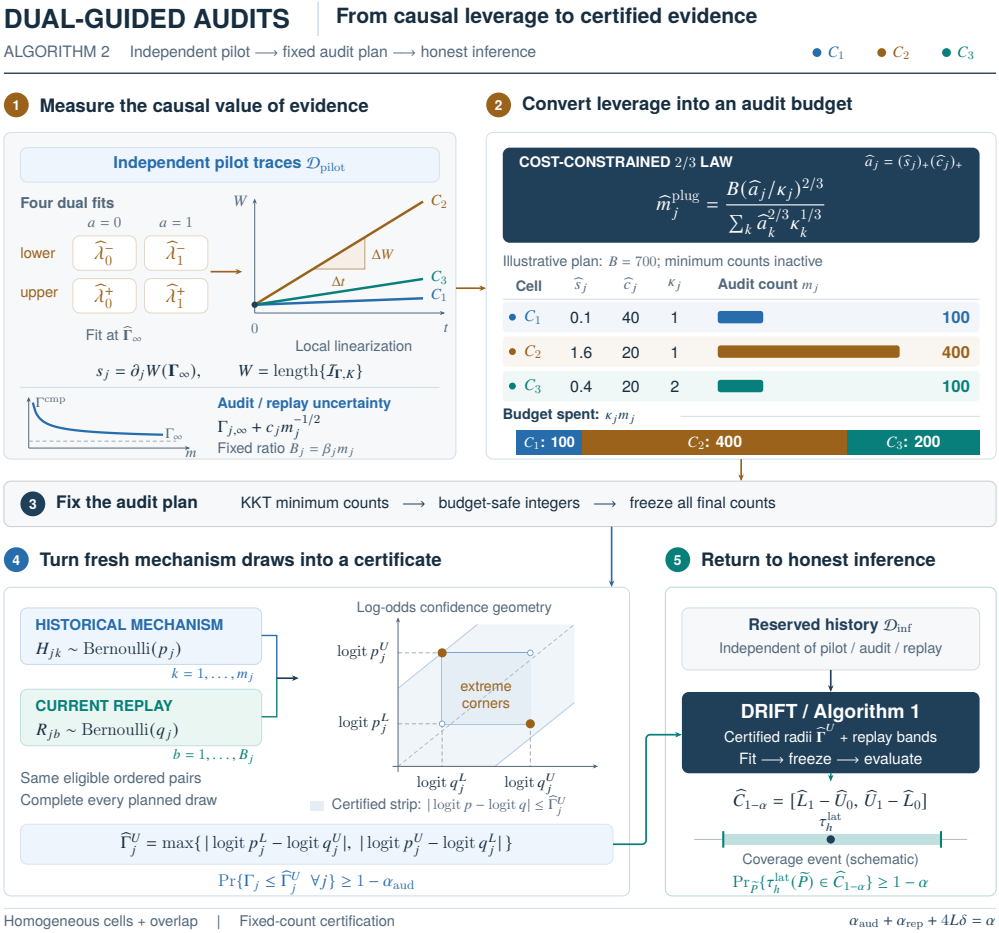

Figure 5: **Historical audits guided by causal-width sensitivity.** An independent pilot estimates leverage and uncertainty; the $2/3$ law fixes counts; fresh audits and replays certify the bridge before DRIFT uses reserved history. Optimality requires homogeneous cells, overlap, regularity, and growing-pilot consistency (Algorithm 2).

τ_h identified set:

$$\inf_{\sum_j \kappa_j m_j \le C_{\text{aud}}} \{W(\mathbf{\Gamma}^{\text{cmp}}(m)) - W(\mathbf{\Gamma}_\infty)\} = \frac{\left(\sum_j (s_j c_j)^{2/3} \kappa_j^{1/3}\right)^{3/2}}{\sqrt{C_{\text{aud}}}} + o(C_{\text{aud}}^{-1/2}), \tag{107}$$

$$W\{\mathbf{\Gamma}^{\text{cmp}}(m^\star)\} - W(\mathbf{\Gamma}_\infty) = \inf_{\sum_j \kappa_j m_j \le C_{\text{aud}}} \{W(\mathbf{\Gamma}^{\text{cmp}}(m)) - W(\mathbf{\Gamma}_\infty)\} + o(C_{\text{aud}}^{-1/2}). \tag{108}$$

Integer rounding changes the two asserted width expansions by $o(C_{\text{aud}}^{-1/2})$.

The oracle allocation depends on unknown s_j and c_j. A growing independent pilot consistently estimates its proportions under the next theorem's conditions.

Theorem E.5 (Plug-in optimality of the audit allocation)**.** *Assume the conditions of Corollary E.4, with fixed J, $\kappa_j > 0$, and $a_j = s_j c_j > 0$ for every cell. Along a sequence of growing pilot samples, let pilot or held-out audit data produce estimates $\widehat{s}_j$ and $\widehat{c}_j$ such that*

$$\widehat{s}_j \xrightarrow{p} s_j, \qquad \widehat{c}_j \xrightarrow{p} c_j, \qquad j = 1, \ldots, J, \tag{109}$$

and let $\widehat{a}_j = (\widehat{s}_j)_+(\widehat{c}_j)_+$, with the allocation defined on the event that all $\widehat{a}_j > 0$. The joint regime is $C_{\text{aud}} \to \infty$ and pilot sample sizes tending to infinity so that (109) *holds. This event has probability*

tending to one. On its complement use any prespecified positive, budget-feasible allocation. A fixed finite pilot does not ensure this consistency. Define the implementable continuous allocation

$$\widehat{m}_j^{\mathrm{plug}} = \frac{C_{\mathrm{aud}}(\widehat{a}_j/\kappa_j)^{2/3}}{\sum_{k=1}^{J} \widehat{a}_k^{2/3}\kappa_k^{1/3}}. \tag{110}$$

Then $\widehat{m}_j^{\mathrm{plug}}/C_{\mathrm{aud}} \xrightarrow{p} m_j^\star/C_{\mathrm{aud}}$ for every j. If integer allocations are obtained by rounding down to preserve the budget (and hence change each count by at most one), then

$$\frac{W\{\mathbf{\Gamma}^{\mathrm{cmp}}(\widehat{m}^{\mathrm{plug}})\} - W(\mathbf{\Gamma}_\infty)}{W\{\mathbf{\Gamma}^{\mathrm{cmp}}(m^\star)\} - W(\mathbf{\Gamma}_\infty)} \xrightarrow{p} 1, \qquad \Gamma_j^{\mathrm{cmp}}(m) = \Gamma_{j,\infty} + c_j m^{-1/2}, \tag{111}$$

where the random allocation is the only plug-in object; the population width still uses the true c_j. Thus the oracle law in (104) *is implementable using consistent pilot estimates; no oracle sensitivity or uncertainty constants are required by the allocation algorithm. With fixed lower bounds $m_j \geq m_{j,0}$, the same conclusion holds for the residual-budget KKT allocation whenever the plug-in active set equals the oracle active set with probability tending to one.*

Uncertainty and causal leverage. Audit value depends on causal leverage and mechanism uncertainty. A rare cell may have large c_j but $s_j \approx 0$ because neither extremal law uses its available weight capacity. A less uncertain cell may determine the lower endpoint. The algebraic $2/3$ exponent follows a standard resource-allocation program; the model-specific coefficient $s_j c_j$ comes from identified-set geometry. The HelpSteer2 radius sweep measures sensitivity to assumed bridge restrictions without collecting audits; it is not an empirical test of the allocation theorem. Figure 3(k,l) reports a separate finite-pilot diagnostic with rankings that vary by baseline and budget.

E.4 How the identified set determines audit value

The derivative s_j depends on which box constraints bind in the extremal reweighting, which calibration moments are expensive, and whether cell j supports the lower endpoint, upper endpoint, both, or neither. The allocation uses the standard convex resource-allocation solution once its coefficients are fixed. Here those coefficients combine the derivative of the specified marginal identified width with the independently certified audit/replay uncertainty rate. The sensitivity model defines the feasible causal laws for eligible pairs; the dual optimizer measures the local value of shrinking that model; and historical mechanism evidence constrains its radius.

Conditional on the independent pilot, final allocations are fixed and fresh audit/replay draws obey (92). Hoeffding's inequality and the union bound hold conditionally with the same error budget; integration over the pilot gives the unconditional certificate. The pilot and final audit/replay information are jointly independent of the historical inference sample, with fixed final counts.

Mechanism audits convert targeted provenance recovery into a certified bridge and allocate evidence where marginal causal width is most sensitive. The HelpSteer2 study examines radius sensitivity and certificate width. The archived EXP8 diagnostic reports fixed-pilot rankings. A nonasymptotic finite-pilot performance guarantee for estimated audit value remains open.

F Experimental Evaluation

Eight synthetic diagnostics address information relocation, identification, two-sample inference, and finite-pilot audit allocation under controlled laws (Section 7). They are analyzed from retained outputs because the historical runner is unavailable. HelpSteer2 adds real generated-response pairs and human-rating proxies under constructed mechanisms and is the primary reproducible experiment. A retained-output Open Bandit diagnostic appears in Section F.4.

F.1 Study design and evaluation scope

We retain $10{,}626$ eligible HelpSteer2 pairs (Wang et al., 2024): the two responses share an exact prompt, have unequal word counts, and have valid original human-rating lists. Ordering shorter/longer defines arms 0/1 without using rating values. A hash-ordered, experiment-specific

split assigns 2,125 pairs to design, 2,125 to development, and 6,376 to confirmation. These are distinct corpus pairs; each simulated historical sample draws n pairs uniformly with replacement from the confirmatory split, with fresh rank, state, and display randomness. Thus n can exceed the number of distinct pairs without adding human ratings.

Helpfulness is the primary outcome and correctness a prespecified check, each divided by four into $[0, 1]$. The target compares mean normalized ratings of complete longer and shorter responses under the fixed corpus and constructed potential-rating law. The responses also differ in content. The study uses real prompts, generated answers, and empirical rating marginals; joint potential-rating coupling, hidden production state, historical randomization, and current replay are constructed.

Ordinary estimators use the selected rating, logged pair/cell identifiers, and independent replay counts. Hidden state, historical propensity, both potential ratings, exact population references, and historical-propensity IPW are reserved for evaluation. Formal effect, coverage, and width analyses began after freezing the design, code, seeds, and prepared-data hashes; earlier whole-data checks concerned structure and rating-list validity. The companion package retains every planned primary condition in a $25{,}000$-row per-repetition table together with numerical diagnostics, source code, configurations, and release checksums. EXP1–8 are supplied as retained data, result tables, and figure assets, not as end-to-end rerunnable experiments.

Table 3: Four-question theory–evidence map for the distinct EXP1–8 and HelpSteer2 designs.

Question	Synthetic studies	HelpSteer2 evidence and boundary
A. Where does information live?	EXP1–3: local contrast, zero drift, and observational equivalence	Hidden association under $\mathbb{E}[p \mid Z] = q$ leaves point bias; strict nonidentification is theoretical
B. Do richer restrictions tighten inference?	EXP4–5: calibration and moment resolution	K, Γ, known-q, and boundary comparisons separate population sets from finite-sample widths
C. How do two samples determine uncertainty?	EXP6–7: joint rate and Gaussian endpoint diagnostic	Complete-set chains, invalid ablations, n/B comparisons, and finite-sample width decomposition
D. Does audit value select the finite-pilot winner?	EXP8: four policies over four budgets	Sensitivity to prescribed bridge radii

F.2 HelpSteer2 evidence across the four questions and boundaries

Controlled mechanisms. Four text cells, fixed using design-split prompt lengths and response-length gaps, have $q = (0.30, 0.45, 0.55, 0.70)$. A shared uniform rank draws both potential ratings from their empirical marginals; its lower/upper half defines hidden state $S = -1/+1$. Historical log-odds differ from q_j by $\eta(0.35v_j + 0.45S)$, where $v = (-1, -1, 1, 1)$ and η denotes experimental drift strength, distinct from the theoretical overlap bound. Thus the maximum absolute shift is 0.8η, equal to 0.48 at the anchor $\eta = 0.6$. The analysis radius Γ is a separate admissible restriction. A second mechanism enforces $\mathbb{E}[p \mid Z] = q$ exactly and varies only the hidden state–rating association. The ordinary information boundary is unchanged across these controls (Section J.2).

Certification protocol and anchor. Two fixed folds, independent replay in all four population cells, simultaneous Clopper–Pearson bands projected into the known range $[0.30, 0.70]$, deterministic score bounds, frozen threshold tie-breaking, zero-dual fallback, and the full empirical-Bernstein correction implement the finite-support specialization. The protocol fixes $\alpha = 0.05$, $\alpha_{\mathrm{rep}} = 0.01$, and eight held-out direction/fold errors of $\delta = 0.005$. Its 15 main, three outcome/coupling, and seven method-variant conditions each have $1{,}000$ completed repetitions.

All 20 positive-radius conditions satisfying the coverage assumptions and both compatible known-q, zero-drift boundary conditions have $1{,}000/1{,}000$ observed chains $\tau \in \mathcal{I} \subseteq \widetilde{\mathcal{I}} \subseteq \mathcal{C}$. The individual two-sided 95% Wilson interval is $[0.9962, 1]$; it measures simulation precision for each condition separately. The positive-radius conditions exercise Theorem D.3; the collapsed zero-drift boxes are checked through Corollary D.4. The observed 100% chain rate summarizes performance on the frozen design; the stated theorem conditions give population coverage.

At $(\eta, \Gamma, n, B) = (0.6, 0.6, 10{,}000, 10{,}000)$, the population interval is $\mathcal{I} = [-0.070, 0.197]$, with width 0.267. Across the $1{,}000$ frozen anchor repetitions, $\widetilde{\mathcal{I}}$ has mean width 0.302 and $\mathcal{C}$ has mean endpoints $[-0.200, 0.322]$, mean width 0.522, and 5–95% width quantiles $[0.506, 0.538]$. Mean endpoints average the random endpoints across repetitions. Every anchor certificate is nonempty and excludes some of the legal $[-1, 1]$ range, but both $\mathcal{I}$ and every final interval in the entire matrix contain zero. The outer restrictions narrow the legal effect range while leaving effect direction unresolved. The endpoint distributions, direct $\mathcal{I} \subseteq \mathcal{C}$ event, and violation magnitudes below are post-hoc descriptive analyses of the frozen records.

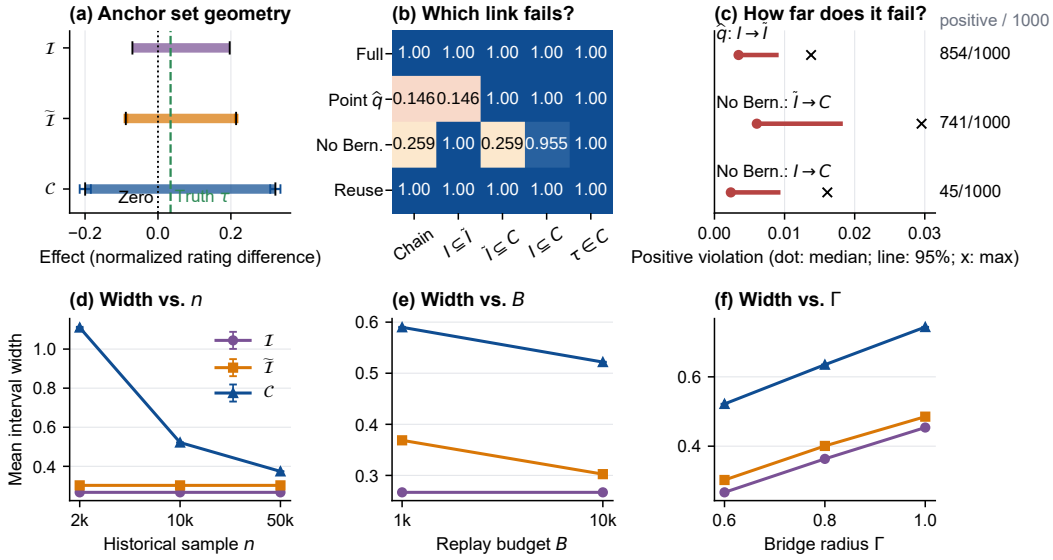

Figure 6: **HelpSteer2 certificate diagnostics.** (a) Bars join mean endpoints; whiskers are endpoint 5–95% quantiles. (b) Link-specific rates distinguish chain from scalar coverage. (c) Positive violation median, 95% quantile, maximum, and count. (d–f) Mean widths with pointwise 95% Monte Carlo intervals. Panels (a–c) and precision summaries describe frozen records post hoc; containment tolerance is 10^{-10}.

Table 4: Frozen main comparisons at $\eta = 0.6$, $K = 4$. Each row has $1{,}000$ complete inclusion chains out of $1{,}000$ repetitions. Widths use normalized rating-difference units; n counts simulated historical traces and B total replay draws.

Comparison	n	B	Γ	$\lvert\mathcal{I}\rvert$	$\lvert\widetilde{\mathcal{I}}\rvert$	$\lvert\mathcal{C}\rvert$
n	2,000	10,000	0.6	0.267	0.303	1.110
Anchor	10,000	10,000	0.6	0.267	0.302	0.522
n	50,000	10,000	0.6	0.267	0.303	0.374
B	10,000	1,000	0.6	0.267	0.369	0.590
Γ	10,000	10,000	0.8	0.363	0.401	0.635
Γ	10,000	10,000	1.0	0.454	0.485	0.744

Question A: where information lives. At fixed $\Gamma = 1$, increasing η from 0 to 1 increases replay-IPW signed mean error from approximately 0.001 to 0.037 and cell-AIPW mean error from approximately zero to 0.048. Figure 7 reports pointwise Monte Carlo intervals rather than taking the absolute value after averaging. Cell-AIPW uses historical cell propensities and outcomes fitted on the complementary fold, so it can adjust observable marginal selection but not the hidden coupling. In the stronger marginal-calibrated control, $\mathbb{E}[p \mid Z] = q$ holds by construction: with hidden association, replay-IPW and cell-AIPW mean errors are 0.047719 (MCSE 0.000679) and 0.047414 (MCSE 0.000212). Without association they are -0.000508 (MCSE 0.000693) and 0.000295 (MCSE 0.000213), and both pointwise intervals include zero (Table 10). This isolates

the contribution of hidden coupling under marginal propensity agreement. Theorem B.2 gives the corresponding strict observational-equivalence construction.

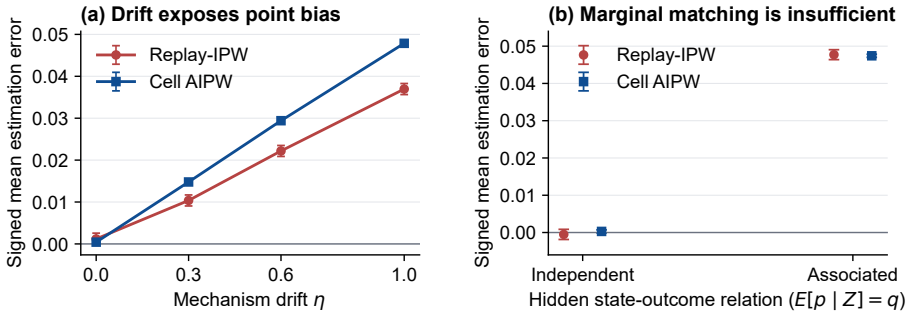

Figure 7: **Marginal selector agreement does not remove hidden-provenance bias.** Left: signed mean error versus constructed drift. Right: under $\mathbb{E}[p \mid Z] = q$, hidden association leaves replay-IPW/cell-AIPW errors $0.048/0.047$; without it, both pointwise MC intervals include zero. Error bars quantify pointwise 95% Monte Carlo uncertainty over $1{,}000$ frozen repetitions.

Question B: richer restrictions and set width. The Γ comparison reuses the same historical/replay bundles at $\eta = 0.6$, so enlarging the assumed radius expands nested sets at a fixed observed-data law. Final mean widths at $\Gamma = 0.6/0.8/1$ are $0.522/0.635/0.744$; relative to the anchor, the paired mean increases are 0.11314 (MCSE 0.00025) and 0.22179 (MCSE 0.00024). By contrast, increasing η changes the historical law; width need not increase when Γ is held fixed. Coarser $K = 1$ calibration gives width 0.550, whereas $K = 2$ and $K = 4$ are nearly equal here. Supplying known q to the DRIFT envelope gives 0.486. At known q and zero drift, both population widths vanish, but the helpfulness/correctness certificates retain widths $0.242/0.246$ and use four expected boundary-fallback components per repetition.

Question C: sampling uncertainty and complete-set certification. At $B = 10{,}000$, increasing n from $2{,}000$ to $10{,}000$ to $50{,}000$ gives final mean widths $1.110/0.522/0.374$; their pointwise 95% Monte Carlo intervals are $[1.107, 1.114]/[0.521, 0.522]/[0.374, 0.374]$. The Bernstein range contribution decreases from 0.532 to 0.106 to 0.021. At the anchor, total width 0.522 is the sum of uncorrected held-out width 0.305, variance correction 0.110, and range correction 0.106, up to rounding. The uncorrected width minus $|\widetilde{\mathcal{I}}|$ is a signed empirical residual reflecting training and evaluation randomness; it is negative in 38.1% of anchor repetitions. Increasing B from $1{,}000$ to $10{,}000$ at $n = 10{,}000$ gives 0.590 to 0.522. The n/B conditions use distinct bundles, whereas method and radius comparisons align shared bundles. Historical concentration is the larger finite-sample cost on this grid. Budget rankings in other designs and the regular-endpoint expansion require separate evaluation.

Ablations test the complete chain. Plug-in-$\widehat{q}$ envelopes and deletion of the Bernstein correction produce literal complete-chain rates 0.146 and 0.259. The first breaks $\mathcal{I} \subseteq \widetilde{\mathcal{I}}$ in $854/1{,}000$ repetitions, with positive violation median/95% quantile/maximum $0.00344/0.00916/0.01378$; nevertheless $\mathcal{I} \subseteq \mathcal{C}$ and $\tau \in \mathcal{C}$ hold in all repetitions. Removing Bernstein breaks $\widetilde{\mathcal{I}} \subseteq \mathcal{C}$ in $741/1{,}000$, with corresponding values $0.00604/0.01831/0.02954$; direct $\mathcal{I} \subseteq \mathcal{C}$ holds in $955/1{,}000$ (pointwise Wilson interval $[0.9403, 0.9662]$), and $\tau \in \mathcal{C}$ still holds throughout. Fitting on the full history and evaluating on its two folds attains every empirical event at 1.000, but violates the required training/evaluation independence. A fixed coarse threshold grid preserves the certificate construction and gives width 0.539 rather than 0.522, illustrating directionally safe approximate training. Reverse-rank coupling and correctness yield chain rates 1.000 and widths $0.515/0.513$. Complete link diagnostics and denominators appear in Tables 8 and 9.

Question D: historical-evidence boundary. The HelpSteer2 radius sweep quantifies sensitivity to assumed provenance restrictions on one corpus with a four-cell selector, discrete ratings, and prespecified synthetic couplings. It does not evaluate the first-order audit-allocation theorem; EXP8 separately reports fixed-pilot allocation under its own synthetic law. Its intervals leave the effect direction unresolved.

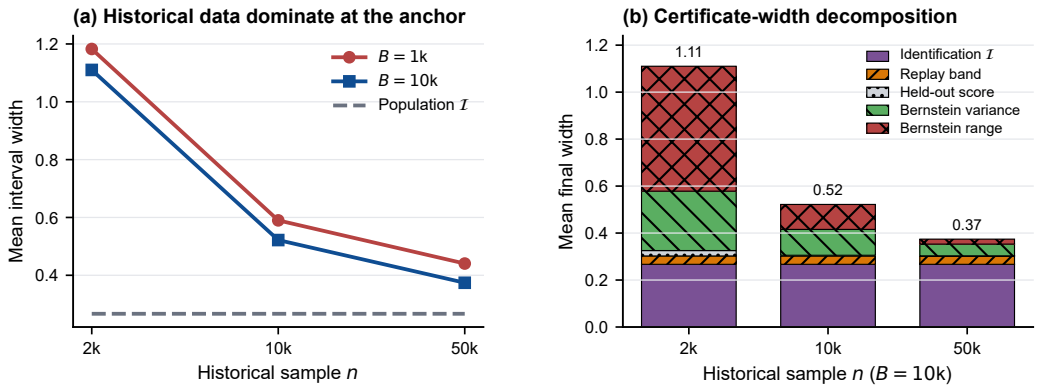

Figure 8: **Sampling precision and fixed-radius ambiguity.** (a) Mean certificate width across the frozen $n \times B$ grid at $\eta = \Gamma = 0.6$. (b) At $B = 10{,}000$, width separates into population identification, replay enlargement, a signed held-out residual, and Bernstein variance/range corrections. The residual is an empirical decomposition component and need not be nonnegative in each repetition.

F.3 Eight archived-output synthetic diagnostics

EXP1–8 retain their original data-generating laws, estimands, numerical summaries, and source panel letters. Their results come from historical runs and are separate from the HelpSteer2 experiment; the companion package supplies retained data, result tables, and figure assets, but not the original runner framework. The groups below follow the four questions in the main text.

Question A: where information lives.

F.3.1 Experiment 1: information relocation

We test the two information laws in Theorem B.1 inside one data-generating process rather than in separately generated armwise and local experiments. In each repetition, DGP-A constructs a single exact-treatment catalog for one conditional stratum, generates n pre-display binary candidate pairs from that catalog, orders each pair using the prespecified step-by-step-explicitness contrast, and randomizes the displayed candidate with a known local design probability bounded away from zero and one. We cross $n \in \{400, 800, 1600, 3200, 6400\}$ with $K_n(\bar{h})/n \in \{0.5, 1, 2\}$ and run 300 independent Monte Carlo repetitions per condition using root seed 20260902. All 4,500 repetitions completed without failure.

The four diagnostics are computed from the same generated-treatment bundle in every repetition. An exact arm is *unseen* if it is never displayed in the historical sample and a *singleton* if it is displayed exactly once. The exact-surface estimator uses each observed arm's sample mean and the observed global outcome mean for unseen arms; its risk is the mean squared error over all $K_n(\bar{h})$ oracle arm means. The local target is the sample mean of the higher-minus-lower conditional outcome contrast in the generated pairs, and the estimator is the Horvitz–Thompson score using the known within-pair design probability. This realized-pair average is a conditional finite-design diagnostic of the localization mechanism, rather than the exact population-MSE target of the theorem. Both estimators are fixed using the observed candidate and outcome trace before oracle quantities are accessed solely for error evaluation.

The occupancy diagnostics show that exact-arm replication does not improve when the support expands proportionally with the sample. Across the five sample sizes, the unseen fraction is approximately 0.141, 0.372, and 0.608 as $K_n(\bar{h})/n$ increases from 0.5 to 1 and 2. The corresponding singleton fractions are approximately 0.271, 0.364, and 0.301, displaying the expected nonmonotone occupancy pattern. These fractions are nearly invariant in n at each fixed support ratio, so increasing the sample does not create repeated information for each exact identity when the catalog grows at the same rate.

The paired risk comparison in Figure 3(a,b) gives the direct diagnostic required by the theorem. Exact-surface risk remains between 0.09504 and 0.11678 over the final grid; its ratio-specific log–log slopes are 0.0020, -0.0002, and -0.0020, and all three 95% intervals contain zero. By contrast, local-contrast RMSE falls from 0.07492, 0.07143, and 0.07254 at $n = 400$ to 0.01744, 0.01719, and 0.01898 at $n = 6400$ for support ratios 0.5, 1, and 2, respectively. The corresponding slopes are -0.5378, -0.5219, and -0.4819, with all three 95% intervals covering the root-n reference -0.5. Thus, within this prespecified finite grid, the full exact-arm surface shows no detectable decay while the local contrast improves at approximately the $n^{-1/2}$ rate.

The synthetic Monte Carlo results support the information-relocation mechanism in Theorem B.1 for the prespecified DGP-A catalog, estimators, and grid. All conditions and repetitions are retained. The fitted slopes summarize behavior over this finite grid.

F.3.2 EXP2: ZERO-DRIFT RECOVERY

We test the exact-replay boundary in Corollary C.8 as a recovery property of the implementation. The prespecified design sets $\Gamma_{\text{true}} = \Gamma_{\text{assumed}} = 0$ and $p_{\text{true}} = q_{\text{true}}$. It crosses $n \in \{400, 800, 1600, 3200\}$ with replay budget $B \in \{600, 1200, 2400\}$ and runs 1,000 repetitions in each of the 12 conditions using root seed 20260905. All 12,000 prespecified repetitions completed; none failed or was replaced.

The ordinary comparator is the LCE plug-in based on $\widehat{q}$. Within each repetition, this estimator, oracle historical inverse-probability weighting (IPW), and the lower and upper DRIFT endpoints are computed from the same simulation bundle. Ordinary procedures receive only the observed replay estimate $\widehat{q}$; the oracle propensity is used only by oracle historical IPW and explicitly labeled equality checks. The reported implementation uses the replay plug-in without an added sensitivity radius, compatibility margin, replay-band inflation, or endpoint widening.

The collapse is exact at the stored numerical precision. Across all conditions, the maximum discrepancy between p_{true} and q_{true} is 1.11×10^{-16}, every DRIFT interval has width exactly zero, and the maximum absolute discrepancy between either endpoint and LCE is 1.46×10^{-16}. The two left panels of row 1 in Figure 11 separate the numerical collapse from the remaining paired Monte Carlo difference to the oracle benchmark: the worst absolute paired z-score is 2.3801, below the prespecified 4-MCSE threshold, with condition-level MCSEs ranging from 4.28×10^{-4} to 9.61×10^{-4}.

These outputs verify algebraic zero-drift plug-in recovery of the population LCE identity. With estimated $\widehat{q}$, collapsed weight boxes can violate empirical calibration moments. Finite-sample coverage, per-repetition equality with oracle IPW, and nonzero-drift behavior require separate checks. The retained EXP2 raw-output file documents the historical computation. The known-q boundary checks in Section F.2 report population collapse and finite-sample certification separately.

F.3.3 EXPERIMENT 3: REPLAY PRECISION WITHOUT HISTORICAL PROVENANCE

We directly instantiate the observational equivalence in Theorem B.2 to test whether increasingly precise current replay can resolve the missing historical assignment mechanism. Each final repetition contains $n = 480$ eligible ordered pairs and one researcher-visible realization of the candidate treatments, historical display D, observed outcome Y, replay draws, and sanitized metadata. That realization is coupled to two oracle latent completions: World 1 has $\tau_h = 1/2$, whereas World 2 has $\tau_h = 1/6$. Only the latent historical state, propensity provenance, potential outcomes, and causal truth differ between the completions; no world label or oracle quantity is available to the ordinary analysis.

The prespecified design sets the current replay probability to $q = 1/2$ and varies only the replay budget over $B \in \{96, 384, 1536, 6144\}$. We retain 200 independently seeded twin pairs at each budget, for 800 prespecified repetitions under root seed 20260908; all 800 completed and none was dropped, replaced, or selectively rerun. Cross-world equality is tested on the exact shared realization with absolute and relative tolerances both equal to zero. As an additional firewall check, the ordinary analysis is rerun for every pair after the oracle and private directories are physically absent, and filenames, columns, metadata, strings, and output schemas are scanned for world or truth leakage.

The atlas comparison in Figure 3(c,d) separates observable equality from replay precision. As B increases from 96 to 6144, MAE falls from 0.03969 to 0.00519 and RMSE falls from 0.04991 to

0.00641; the intermediate RMSE values are 0.02331 and 0.01215 at $B = 384$ and 1536. The products $\mathrm{RMSE}(\widehat{q})\sqrt{B}$ are 0.489, 0.457, 0.476, and 0.503, respectively, consistent with the descriptive Bernoulli $B^{-1/2}$ sampling scale. Replay precision is assessed descriptively, with all budget conditions retained regardless of fitted slope or finite-sample monotonicity.

Every twin pair has zero observed mismatch, maximum numerical difference zero, and zero world-identity, oracle-column, or oracle-path leakage; the isolated observed-only rerun passes for all 800 pairs. Nevertheless, the same figure shows that the two oracle causal truths remain separated by $1/3$ at every replay budget, exceeding the prespecified 0.30 threshold. Hence arbitrarily precise information about the current q does not recover the association between the historical production state, the historical display probability, and potential outcomes. The experiment isolates nonidentification from current replay alone. A valid bridge or historical audit supplies the additional provenance information needed for causal analysis.

Implementation versus population coverage. The fixed six-stratum replay generators use independent random streams and predetermined allocation, supporting all-cell replay bands. Experiment 4 nevertheless chooses its feasibility adjustment from the full historical sample and uses a same-sample linear program; its observed containment is an empirical diagnostic, not an implementation of the cross-fitted population theorem. Experiment 5 uses sample splitting and a fixed drift radius, but its code infers the stratum count from observed historical strata and evaluates some bounds only on observed contexts. Its implementation therefore needs an additional all-population-cell check before invoking Theorem D.3, particularly when a positive-mass stratum is absent from the historical sample. The reported outputs follow these prespecified procedures and their stated coverage conditions.

Question B: richer restrictions and inference.

F.3.4 Experiment 4: robustness to historical–current drift

We evaluate the nonzero-drift identified set of Definition C.5 in a finite-stratum generated-pair design. Every record is an eligible ordered binary pair. Six fixed context strata govern both the current replay probability q_j and the historical display probability p_j; their log-odds difference has alternating direction and maximum radius Γ. The historical propensity depends on the stratum but not on the hidden outcome state. The latter affects potential outcomes and is withheld from every ordinary procedure. Full formulas and the outcome-blind feasibility rule appear in Section J.3.

The prespecified grid is $\Gamma \in \{0, 0.15, 0.30, 0.45, 0.60\}$. At each value we retain 400 independent repetitions with $n = 600$ eligible logged pairs and an independent stratified replay budget $B = 3000$, for 2,000 repetitions and 10,000 method rows in total. The five procedures are Historical oracle IPW (Oracle Benchmark Only), Current replay IPW, replay augmented inverse-probability weighting (AIPW), box-only sensitivity, and six-stratum calibrated Drift. Replay AIPW uses arm-specific ridge linear outcome regressions, clipped to $[0.01, 0.99]$, together with the observed replay estimate. The two sensitivity procedures begin with the same simultaneous replay band and certified radius $\Gamma+0.08$; calibrated Drift then uses the minimum additional outcome-blind inflation required for empirical arm-by-stratum normalization. No failed replication was removed or replaced.

Table 5 separates point-interval behavior from set containment. The implemented outer intervals from both sensitivity procedures contain the replication-specific causal truth in every repetition at all five drift values. Across the point procedures, empirical containment ranges from 0.865 to 0.920 and is not monotone in Γ; the relative containment rates vary across the grid. The simultaneous true-q replay-band event holds in 1962 of 2000 repetitions, exceeding the prespecified finite-Monte Carlo cutoff of 1949. This meets the finite-Monte Carlo criterion for assessing nominal simultaneous coverage of 0.98. Separately, the oracle-firewall integrity audit passes and the true historical propensity lies in the final envelope in all 2000 repetitions.

The population and finite-sample results in Table 5 show the cost of drift and the gain from valid moment information. As Γ increases from 0 to 0.60, box-only population identified width rises from 0 to 0.376653, whereas calibrated width rises from 0 to 0.101692. The corresponding implemented outer widths rise from 1.937195 to 3.259325 and from 1.866179 to 3.041807. Thus six-stratum

Table 5: Set-estimation results in the prespecified drift experiment. Containment is the fraction of 400 repetitions whose implemented outer interval contains the replication-specific causal truth. Population identified width, implemented finite-sample outer width, and Hausdorff endpoint error are distinct metrics.

Γ	Method	Containment	Identified width	Outer width	Hausdorff error
0.00	Box-only	1.000	0.000000	1.937195	1.002073
	Calibrated DRIFT	1.000	0.000000	1.866179	0.964884
0.15	Box-only	1.000	0.093123	2.154023	1.066650
	Calibrated DRIFT	1.000	0.025925	2.069917	1.055158
0.30	Box-only	1.000	0.191890	2.473841	1.171865
	Calibrated DRIFT	1.000	0.051834	2.348405	1.177804
0.45	Box-only	1.000	0.288544	2.870120	1.322624
	Calibrated DRIFT	1.000	0.077061	2.696904	1.342791
0.60	Box-only	1.000	0.376653	3.259325	1.471506
	Calibrated DRIFT	1.000	0.101692	3.041807	1.500209

calibration contracts both width measures throughout this prespecified grid. The larger outer widths include the implemented full-sample symmetric finite-sample expansion.

The Hausdorff-error columns of Table 5 show the associated endpoint-accuracy tradeoff. Mean Hausdorff endpoint error rises from 1.002073 to 1.471506 for box-only sensitivity and from 0.964884 to 1.500209 for calibrated DRIFT; calibrated endpoint error is not smaller at every nonzero drift value. The procedures exhibit a tradeoff among containment, width, and endpoint accuracy. All Monte Carlo precision and four-block stability checks pass. Outcome-blind feasibility inflation is reported as a compatibility diagnostic in Section J.3.

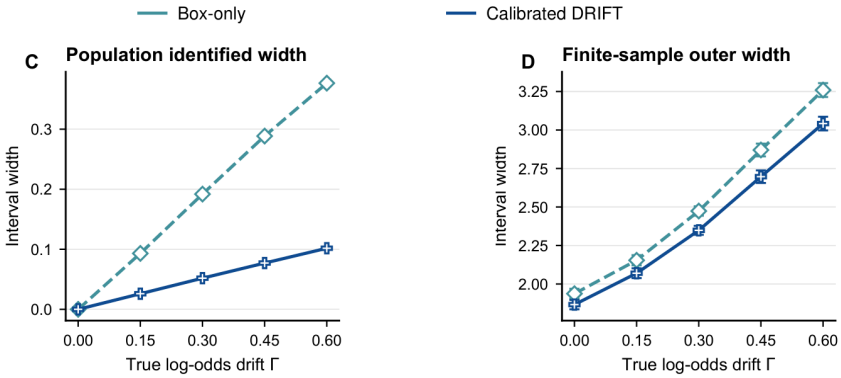

Figure 9: **EXP4 — Calibration and finite-sample width.** Calibration contracts both population identified width and implemented outer width on the tested drift grid ($n = 600$, $B = 3000$, 400 repetitions per condition). The outer-width bars are pointwise 95% Monte Carlo intervals; outer widths are not population identified widths, and endpoint-error rankings are not uniform.

F.3.5 EXPERIMENT 5: MOMENT RESOLUTION AND THE IDENTIFICATION–UNCERTAINTY TRADEOFF

We test the nestedness prediction of Theorem C.10 while keeping population identification and finite-sample certification conceptually separate. The former asks whether valid additional moments remove marginal trace weights and weakly contract the population causal interval. The latter is the cross-fitted implementation motivated by Theorem D.3, whose width also includes replay-band and held-out empirical-Bernstein corrections and therefore need not decrease with sieve dimension. Population coverage for this implementation remains subject to the all-population-cell qualification above.

The prespecified design uses the six-stratum generated-pair law at $\Gamma = 0.6$, with $n = 600$ eligible logged pairs, a stratified replay budget $B = 3000$, and 400 independent repetitions. Within each repetition the same observed bundle is evaluated at $K \in \{1, 3, 5, 6\}$. Starting from an intercept, the nested basis adds the observed continuous context and normalized stratum index, their squares, and

finally their interaction. Two row-index folds are fixed before outcome access. The sieve grid, basis, seed schedule, solver tolerances, and normal-versus-conservative routing rule were fixed before the final results were observed; all 1,600 repetition–K rows were retained. Further implementation details appear in Section J.4.

The paired panels in Figure 10 report the population and finite-sample quantities on separate scales.

Population intervals contract from $[-0.065638, 0.311015]$ at $K = 1$ to $[-0.052519, 0.309336]$ at $K = 3$. At $K = 5$, the interval is $[-0.040895, 0.307442]$; the $K = 6$ endpoints agree to the reported precision. The corresponding widths are 0.376653, 0.361855, 0.348338, and 0.348338. Thus the population sets are weakly nested, with no additional contraction from the final basis term.

Finite-sample behavior differs from the population contraction. Mean outer widths (Monte Carlo standard errors) are 5.312477 (0.026359), 6.365559 (0.031169), 8.346701 (0.090959), and 9.893869 (0.246816) for $K = 1, 3, 5, 6$, respectively. The smallest observed mean on the prespecified grid is therefore at $K = 1$, not at the largest sieve. Paired mean increments for $1 \to 3$, $3 \to 5$, and $5 \to 6$ are 1.053082, 1.981142, and 1.547168; their 95% Monte Carlo intervals are $[1.018710, 1.087455]$, $[1.825770, 2.136513]$, and $[1.132783, 1.961554]$. These Monte Carlo intervals quantify simulation precision for the contrasts on the specified grid.

The routing audit distinguishes conservative certification from solver error. Of 6,400 fold–arm–K systems, 6,380 use optimized dual certificates and 20 solver-certified infeasible empirical systems, occurring across 13 repetitions, use the prespecified zero-dual certificate. No unknown, unbounded, numerical-error, exception, or inconsistent path occurs. The fallback components address empirical training infeasibility while retaining the corresponding repetitions; population feasibility is a separate condition. The experiment verifies marginal trace-model nestedness and measures the finite-sample concentration cost of richer valid moments at this fixed DGP, n, and B.

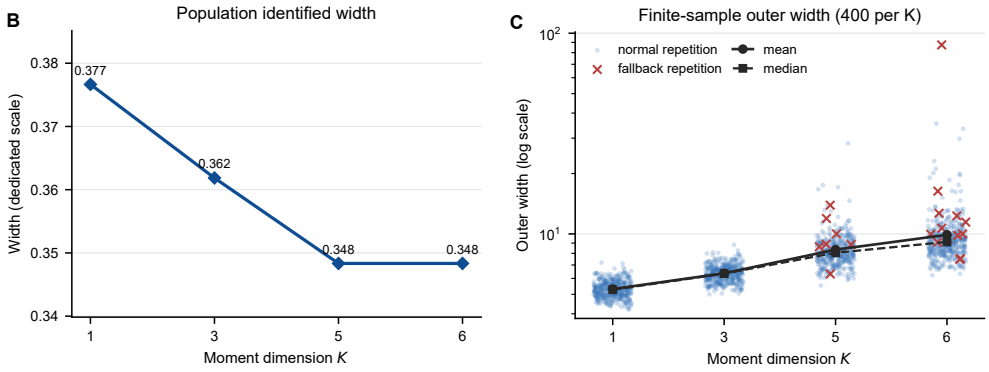

Figure 10: **EXP5 — Identification versus certification.** Population width contracts with moment resolution, whereas finite-sample outer widths increase on this grid. All 400 repetitions per K, including fallbacks, are retained.

Question C: historical and replay uncertainty.

F.3.6 EXP6: JOINT HISTORICAL-REPLAY RATE

We examine a descriptive finite-grid fit motivated by the regular two-sample expansion in Theorem D.7, varying the eligible historical sample size n and independent replay budget B jointly. The reported estimator applies an arm-order projection before forming the effect endpoints. The target is the lower or upper population marginal-model endpoint at fixed $K = 3$ and $\Gamma = 0.6$. The prespecified finite-grid summary model is

$$\mathrm{RMSE}^2\{\widehat{\theta}_e(n, B)\} \approx \frac{C_{1,e}}{n} + \frac{C_{2,e}}{B}, \qquad e \in \{L, U\}, \tag{112}$$

fitted without an intercept by nonnegative least squares to the projected-endpoint MSE. This joint surface, rather than separate marginal log–log regressions, is the primary empirical object.

The prespecified six-stratum design crosses $n \in \{1200, 2400, 4800\}$ with $B \in \{240, 480, 960, 1920\}$. Each of the 12 cells contains 400 independent repetitions, for 4,800 completed repetitions with no retry, replacement, or result-contingent tuning. The bounded continuous outcome law preserves the conditional means of the finite-stratum design while avoiding outcome atoms at the dual kinks. Replay counts use the globally fixed continuity correction $\widetilde{q}_j = (S_j + 1/2)/(B_j + 1)$, whose $O(B^{-1})$ perturbation is asymptotically negligible relative to $B^{-1/2}$. Two row-index folds and the three-term basis $(1, X_2, 2j/5 - 1)$ are fixed before outcomes are accessed; complete implementation details appear in Section J.5.

Table 6: Joint-rate fits for the arm-order-projected estimators of the two population endpoints. The C_2 intervals are grouped-bootstrap 95% intervals over 10,000 resamples; R^2 is descriptive on the MSE scale.

Endpoint	$\widehat{C}_1$	$\widehat{C}_2$	C_2 interval	R^2	Prespecified theory (C_1, C_2)
Lower	5.1741	0.6999	$[0.2745, 1.1959]$	0.5102	$(0.3111, 0.0576)$
Upper	5.3952	0.7381	$[0.2891, 1.2385]$	0.5280	$(0.3162, 0.0459)$

The endpoint heatmaps in Figure 3(i,j) and coefficients in Table 6 use the arm-order-projected estimates: whenever an arm's lower estimate exceeds its upper estimate, both are replaced by their midpoint before combining the arms. This occurs in $132/4{,}800$ repetitions (2.75%), with maximum absolute effect-endpoint change 4.2771. The projected fits have positive historical and replay coefficients; unconstrained no-intercept fits return the same coefficients, and bootstrap boundary frequencies for $\widehat{C}_2$ are 0.000 and 0.001 for the lower and upper endpoints. A post-hoc comparison using the retained unprojected arm endpoints and the same grid-level MSE fit gives $(\widehat{C}_1, \widehat{C}_2) = (0, 24.2445)$ for the lower endpoint and $(0, 24.9201)$ for the upper endpoint. The reported rate fit is therefore sensitive to the projection.

A direct replay diagnostic isolates the second sampling scale. Across the four replay budgets, raw replay-probability RMSE multiplied by $\sqrt{B}$ stays between 1.187 and 1.199, and its log–log slope is -0.5002; the continuity-corrected slope is -0.4903. These values support $B^{-1/2}$ scaling for replay-probability error. Endpoint RMSE also depends on historical sample size and optimization.

Both population benchmark pairs in Table 6 are materially smaller than the projected empirical coefficients. The larger coefficients, moderate descriptive R^2 values, and projection sensitivity limit this grid to a descriptive diagnostic; it does not directly validate the regular expansion for the unprojected estimator. All solver checks succeed, but this design does not isolate higher-order, dual-learning, and projection effects.

F.3.7 Experiment 7: nonzero-drift optimized endpoint inference

We next evaluate the Gaussian approximation for the *optimized* partial-identification endpoints rather than for a fixed IPW score. The prespecified design sets $\Gamma = 0.35$, $K = 3$, two balanced cross-fitting folds, and independent historical and finite-stratum replay samples with $n = B = 21{,}120$. The bounded continuous outcome law preserves the finite-stratum propensities, masses, and conditional means while placing zero probability on population-dual kinks. The lower and upper population endpoints are 0.0577685 and 0.1862273. All 1,000 repetitions were fixed in advance and completed without replacement or deletion; the two-sample standard error combines the joint historical score covariance with the independent replay contribution as in (83). Full protocol and implementation checks appear in Section J.6.

The lower- and upper-endpoint QQ panels and distributions in Row 2 of Figure 11 support a useful Gaussian approximation with visible finite-sample qualifications. The lower endpoint has bias -0.000515, empirical SD 0.004257, mean estimated SE 0.004273, and studentized mean/SD $-0.142/1.001$; its nominal 95% interval covers in 95.7% of repetitions (Monte Carlo SE 0.64 percentage points). The upper endpoint has bias 0.000570, empirical SD 0.004382, mean estimated SE 0.004218, and studentized mean/SD $0.161/1.038$; its nominal 95% Wald interval covers the upper endpoint in 93.2% of repetitions (Monte Carlo SE 0.80 percentage points). Central Q–Q regions are approximately linear and both histograms are bell-shaped, but lower-left and upper-right

tail deviations remain. Relative to $N(0,1)$ calibration, the upper endpoint shows mild rightward displacement, scale inflation, and undercoverage.

All $4{,}000$ training fold–arm feasibility programs are compatible, all $8{,}000$ primary dual fits are optimal, and every independent numerical cross-check succeeds. No repetition has a nonfinite or reversed endpoint, nonpositive standard error, radius hit, or optimizer failure. Active-branch disagreement is 5.149% and serves as a stability diagnostic. The small biases exceed Monte Carlo noise; higher-order dual-learning and replay-adaptation effects are possible contributors, though this design does not separate them. No bias correction or result-contingent adjustment is applied. The 93.2% rate concerns Gaussian endpoint calibration. Scalar-effect and set-chain coverage are addressed by the separate outer-certificate argument in Theorem D.3.

Question D: finite-pilot audit allocation.

F.3.8 Experiment 8: finite-pilot mechanism-audit allocation

We compare four implementable audit policies under the same six-stratum binary-outcome law at $\Gamma = 0.6$, with $n = 12{,}000$ eligible logged pairs and $J = K = 6$. The final design fixes 400 new independent repetitions under seed 2026090582. Every repetition includes a finite pilot of $4{,}000$ historical-mechanism draws per cell, used to estimate the empirical width sensitivities. Uniform, Uncertainty-only, Leverage-only, and Full plug-in share the history and pilot and use common prefixes of an independent main audit/replay pool. Uniform allocates equal cost, not equal counts. The four main budgets are $12{,}000$, $24{,}000$, $48{,}000$, and $96{,}000$; actual total cost includes the pilot cost $42{,}800$ and integerized main allocations. Replay has zero incremental cost. Section J.7 gives the prespecified allocation rules, certification, and complete results.

Conditional mean widths decrease across the tested budgets for every policy. Full plug-in falls from 0.441890 (MCSE 0.000917) to 0.192257 (0.000980), while Uniform falls from 0.438179 (0.000893) to 0.190837 (0.001012). Of $6{,}400$ recorded outcomes, $6{,}390$ are feasible and 10 are incompatible, with zero program or solver errors. All incompatibilities occur at main budget $96{,}000$, in four distinct repetitions; failures within a repetition can be correlated. Because some widths are undefined there, the all-400 unconditional mean is unavailable. Costs incurred by every repetition and the feasible-subset means are both retained in Table 11.

The width–cost and paired-difference panels in Figure 3(k,l) show that Full plug-in performance varies by budget and baseline in the jointly feasible repetitions of this six-stratum DGP. Its Full-minus-Uniform differences at the four increasing budgets are 0.003712, 0.000928, -0.000117, and 0.001263; the pointwise Monte Carlo intervals exclude zero in the positive direction at the first, second, and fourth budgets, while the third is unresolved. Full is narrower than Leverage-only at all four budgets, with differences from -0.002023 to -0.000722. Relative to Uncertainty-only, Full is wider at $12{,}000$, narrower at $48{,}000$, and unresolved at the other two budgets. Table 12 reports all 12 paired effects, MCSEs, and 95% intervals using jointly feasible repetitions, rather than subtracting means from different feasible subsets. The intervals are pointwise and unadjusted for multiplicity; comparisons whose intervals include zero remain unresolved.

The allocation law in Theorem E.3 and Corollary E.4 optimizes a population first-order objective. The two right panels of row 1 in Figure 11 support a local expansion for this DGP and show an approximately 0.335% gain for that objective. Finite-pilot empirical intervals address a separate question; applying Theorem E.5 requires consistent pilot sensitivities at the irreducible radius. A nonasymptotic finite-pilot excess-width or regret bound would need to account for sensitivity error, pilot cost, constrained integer allocation, random certified radii, and possible infeasibility.

EXP6–8 show a projection-sensitive descriptive endpoint fit, mild undercoverage for one Gaussian endpoint interval, and finite-pilot audit rankings that vary by budget and baseline. Each finding concerns a refinement beyond the finite-sample outer certificate.

F.4 Supplementary fixed-action Open Bandit diagnostic

The Open Bandit Dataset provides a fixed-action setting for testing the mechanism-provenance calculation under production reward and propensity geometry. It contains recommendation logs from ZOZOTOWN and data from a randomized policy on the same action space (Saito et al., 2021). We

use $10{,}000$ Bernoulli Thompson-sampling interactions and $10{,}000$ random-policy interactions over 80 items. The randomized log supplies an observed uniform-policy click-through benchmark of 0.0038. The production log includes its actual propensity p, which is reserved for oracle diagnostics and simulated mechanism audits.

The public data provide fixed-action logs, with pre-display binary pairs and archived production versions unavailable. The stress test retains the logged contexts, actions, rewards, and production propensities and constructs the replay score as

$$q_i = \mathrm{clip}\left[p_i \exp\{0.45\tanh(Z_i) + 0.30\sin(2\pi T_i) - 0.12(\mathrm{position}_i - 2)\}\right], \tag{113}$$

where Z_i is a standardized item feature and T_i normalized time. The stated transformation imposes controlled mismatch. User strata provide calibration moments, and the target policy is uniform over items. The diagnostic evaluates the finite-action replay envelope, benchmark containment, and audits under the logged reward and propensity geometry. Identification of τ_h additionally requires a pre-display binary-pair trace.

On the full log, oracle production IPW is 0.00236, replay IPW 0.00229, and self-normalized replay 0.00199, all below the randomized benchmark. The gap cannot be attributed entirely to injected drift: oracle IPW has effective sample size only 340 of $10{,}000$, so ordinary importance-weight variance is substantial. Replay mismatch and weak overlap are distinct failure modes.

The raw auxiliary DRIFT set is $[0.00091, 0.00574]$ and contains the randomized benchmark.

The sampling-expanded diagnostic interval is $[0.00026, 0.00985]$. Across 180 subsamples of size $6{,}500$, it contains the benchmark in 98.9% of repetitions. Row 3 of Figure 11 separates sampling uncertainty, audit contraction, bias–variance geometry, and the operating summary. Point estimates omit mechanism uncertainty. Budgeted audits of known production propensities contract the auxiliary set while preserving benchmark containment; outcome noise and overlap still contribute to width. This controlled fixed-action calculation evaluates the replay envelope and audit mechanism against the randomized-log benchmark. The population certificate theorem applies to the separate generated-pair experiment.

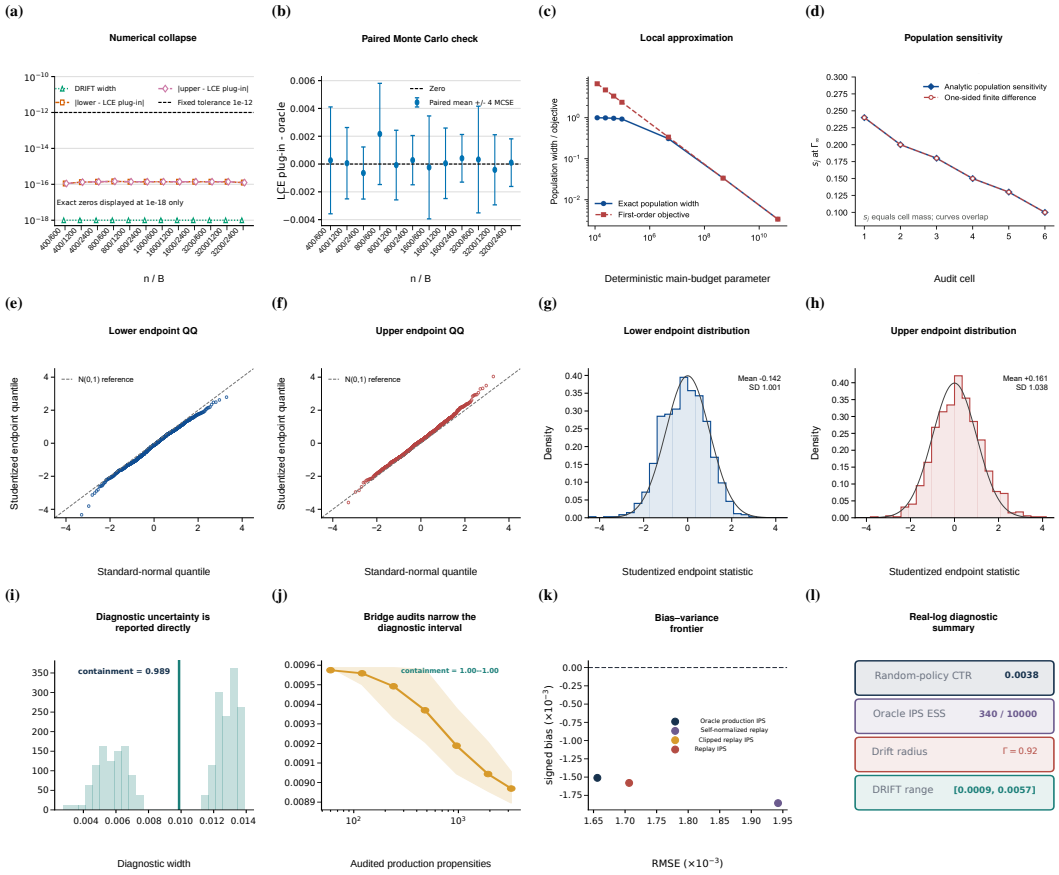

Figure 11: Supplementary boundary and fixed-action diagnostics. Panels (a)–(l) run left to right across rows. Row 1 shows EXP2 zero-drift recovery in (a,b) and the EXP8 population first-order audit diagnostic in (c,d). Row 2 evaluates Gaussian endpoint inference in EXP7: the central distributions are approximately normal, with mild upper-tail displacement and 93.2% upper-endpoint Wald coverage. The finite-sample outer certificate uses a separate argument and is evaluated under a different DGP. Row 3 reports real-log uncertainty, audit contraction, bias–variance geometry, and the operating summary; Open Bandit evaluates fixed-action inference under injected mismatch.

G Discussion and Conclusion

Replay precision and historical provenance. Historical outcomes and current replay answer different questions. Replay estimates how the available selector behaves on saved candidate pairs. Historical mechanism evidence determines how those probabilities constrain the assignment law that generated the outcomes. DRIFT connects the two through a provenance bridge, observable calibration, held-out dual certificates, and derivatives that value additional evidence. Precise knowledge of the current mechanism can still leave a historical causal effect partially identified.

Identification, inference, and audits. Randomized local comparisons reduce dependence on repeated exact-treatment identities. The local IPW estimator requires historical assignment probabilities. Under a fixed or independently audited bridge, DRIFT yields an interval sharp for the specified marginal trace-weight model and outer-valid for compatible latent historical effects. Weak duality preserves the certificate direction under approximate optimization. Train–freeze–evaluate sample splitting and population-valid replay bands give finite-sample outer confidence sets. Regular endpoints admit a joint historical/replay expansion; nonregular endpoints can have non-Gaussian limits. The audit rule allocates evidence according to causal provenance leverage, uncertainty reduction, and cost.

Target and design. The target averages the higher-versus-lower effect over eligible pairs produced by the historical generator. It compares complete generated answers, excludes pairs tied under the prespecified contrast, and requires a fixed pair before randomized final display. A different generator changes the target population, and an outcome-adaptive contrast requires selection-aware inference. The theory assumes independent snapshots, bounded outcomes, local positivity, and a binary slate; deterministic selection, unmodeled overrides, and post-display matching require other designs. Extensions include adaptive dependence, interference-aware local experiments, extraction from larger slates, and transport across evolving generators.

Empirical findings and open questions. The retained EXP1–3 outputs separate local information from historical provenance. EXP4–5 distinguish population contraction from finite-sample certificate width, and EXP6–7 show finite-grid remainders and mild Gaussian upper-endpoint undercoverage. EXP8 reports finite-pilot audit rankings that vary by budget and baseline; its archived summaries have partial raw support, while the plug-in theorem requires a growing consistent pilot. These diagnostics are not end-to-end rerunnable from the release. The HelpSteer2 result table provides the primary reproducible evidence: complete-set inclusion holds in 20 positive-radius conditions and two zero-drift boundary checks, while point-replay and no-Bernstein ablations break set links even when scalar truths remain covered. The Open Bandit retained outputs provide a separate fixed-action stress test under injected mismatch. Independent full-run reconstruction and a nonasymptotic finite-pilot performance guarantee remain open.

Empirical interpretation. The HelpSteer2 study preserves real generated responses and rating marginals while constructing their joint potential outcomes and both selection mechanisms. Its $6{,}376$ confirmatory pairs form a finite corpus sampled with replacement. Anchor certificates exclude part of the legal effect range, but every final interval contains zero, leaving the direction of the effect unresolved. A second corpus and naturally logged, versioned historical randomization would test external validity beyond this controlled construction.

Preserving causal evidence. Generated systems should archive the provenance needed to reconstruct exposure selection: versioned selector probabilities, candidate-set metadata, override logic, and controlled overlap traffic, alongside treatments and outcomes. When complete archival is too costly, dual-guided audits can prioritize historical evidence. This connects partial identification with sequential evidence acquisition and suggests that adaptive systems preserve mechanism information needed for later causal analysis.

H PROOFS OF THE MAIN RESULTS

This appendix proves the main theoretical results; the optimal-transport and knapsack proofs follow their statements in later appendices. Equalities and inequalities hold up to the relevant almost-sure equivalence classes. Local-trace expectations use the law conditional on the eligible ordered-pair event $E = 1$; exact-arm support arguments condition on a fixed history value $\bar{h}$. For a fixed arm a, write $I = I_a$, $\ell = \ell_a$, $u = u_a$, $\phi = \boldsymbol{\phi}_K$, and $m = \mathbb{E}\phi(Z)$ when arm indices are suppressed. In those proofs, $\lambda \in \mathbb{R}^K$ denotes the vector dual variable $\boldsymbol{\lambda}$ used in the main text; scalar multipliers are identified separately.

H.1 PRELIMINARY MEASURE-THEORETIC FACTS

Lemma H.1 (Weak-$\star$ compactness of a calibrated order interval)**.** *Let P be a probability measure, let $\ell, u \in L^\infty(P)$ satisfy $\ell \le u$, let $b_1, \dots, b_K, c \in L^1(P)$, and let $d \in \mathbb{R}^K$. Then*

$$\mathcal{C} = \{w \in L^\infty(P) : \ell \le w \le u,\ \mathbb{E}(wb_k) = d_k,\ k = 1, \dots, K\}$$

is convex and weak-$\star$ compact. The map $w \mapsto \mathbb{E}(wc)$ is weak-$\star$ continuous and therefore attains its minimum and maximum on every nonempty $\mathcal{C}$.

Proof. Convexity is immediate. Put $R = \max(\|\ell\|_\infty, \|u\|_\infty)$. The order interval $[\ell, u]$ is contained in the closed radius-R ball of $L^\infty(P)$, which is weak-$\star$ compact by Banach–Alaoglu because $L^\infty(P)$ is the dual of $L^1(P)$ after the usual identification on a probability space.

It remains to show weak-$\star$ closedness. If $w_\nu \to w$ weak-$\star$ and $w_\nu \ge \ell$, then for every nonnegative $g \in L^1(P)$, $\mathbb{E}\{g(w - \ell)\} = \lim_\nu \mathbb{E}\{g(w_\nu - \ell)\} \ge 0$. Taking $g = \mathbf{1}\{w < \ell - \varepsilon\}$ and then letting $\varepsilon \downarrow 0$ yields $w \ge \ell$ almost surely. The proof of $w \le u$ is identical. Each equality constraint is weak-$\star$ closed because $b_k \in L^1(P)$. Hence $\mathcal{C}$ is a weak-$\star$ closed subset of a weak-$\star$ compact set. Finally, $w \mapsto \mathbb{E}(wc)$ is precisely evaluation of the dual element w at $c \in L^1(P)$, so it is weak-$\star$ continuous. The extreme-value conclusion follows. □

Lemma H.2 (A uniform nondegeneracy constant)**.** *Suppose $\|\phi(Z)\|_2 \le C_K$ almost surely and $G_a = \mathbb{E}\{I_a \phi(Z)\phi(Z)^\top\}$ is positive definite. Then*

$$\kappa_a := \inf_{\|v\|_2 = 1} \mathbb{E}\{I_a |v^\top \phi(Z)|\} > 0.$$

Proof. Let $\lambda_{\min}(G_a) > 0$. For every unit vector v, boundedness gives

$$\mathbb{E}\{I_a (v^\top \phi)^2\} \le \|v^\top \phi\|_\infty \mathbb{E}\{I_a |v^\top \phi|\} \le C_K \mathbb{E}\{I_a |v^\top \phi|\}.$$

The left side equals $v^\top G_a v \ge \lambda_{\min}(G_a)$. Therefore $\mathbb{E}\{I_a |v^\top \phi|\} \ge \lambda_{\min}(G_a)/C_K$, uniformly over the unit sphere. □

H.2 EXPANDING-ARM AND LOCAL-RANDOMIZATION RESULTS

Proof of Theorem B.1. We first prove the exact-arm claim on the stated no-sharing submodel. Under $p_{i,n} = 1/2$, symmetry of the uniformly generated pair gives

$$\mathbb{P}(A_{i,n} = a) = \frac{1}{K_n}, \qquad a \in \mathcal{T}_n.$$

Fix $\delta_n = \min\{1/8, \sqrt{K_n/n}/16\}$ and restrict further to the hypercube $\mu^\omega_{a,n} = 1/2 + \delta_n \omega_a$, $\omega \in \{-1, 1\}^{K_n}$. Let $\omega^{(j)}$ reverse the jth sign. The candidate-pair and display laws are identical under the two adjacent parameters. The only difference in one observation occurs when the displayed exact treatment is j, an event of probability $1/K_n$; conditional on that event the two outcome laws are Bernoulli$(1/2 + \delta_n)$ and Bernoulli$(1/2 - \delta_n)$. For $\delta \le 1/8$, the elementary Bernoulli bound

$$\mathrm{KL}(\mathrm{Bern}(1/2 + \delta) \,\|\, \mathrm{Bern}(1/2 - \delta)) \le 16\delta^2$$

therefore gives

$$\mathrm{KL}(P_\omega, P_{\omega^{(j)}}) \le \frac{16 n \delta_n^2}{K_n} \le \frac{1}{16}.$$

Pinsker's inequality yields total variation at most $1/4$. For any estimator based on the full trace $\mathcal{O}_n$, define $\widehat{\omega}_j = 1$ when $\widehat{\mu}_{j,n} \geq 1/2$ and -1 otherwise. A sign error incurs squared error at least δ_n^2, so the usual hypercube testing argument gives

$$\sup_{\mu_n} \mathbb{E}_{\mu_n}\left[\frac{1}{K_n}\sum_{a=1}^{K_n}(\widehat{\mu}_{a,n} - \mu_{a,n})^2\right] \geq \frac{3\delta_n^2}{8}.$$

Since $\delta_n^2 \geq \frac{1}{256}\min\{1, K_n/n\}$, this proves Part (i), for example with $c_0 = 3/2048$.

For Part (ii), set

$$\psi_{i,n} = \frac{D_{i,n}Y_{i,n}}{p_{i,n}} - \frac{(1-D_{i,n})Y_{i,n}}{1-p_{i,n}}.$$

Consistency and the conditional Bernoulli draw imply

$$\mathbb{E}(\psi_{i,n} \mid \mathcal{G}_{i,n}) = Y_{i,n}^H - Y_{i,n}^L,$$

so $\mathbb{E}\widehat{\tau}_{\mathrm{local},n} = \tau_n$. Because the units are independent and $0 \leq Y_{i,n}^H, Y_{i,n}^L \leq 1$,

$$\mathbb{E}(\psi_{i,n}^2 \mid \mathcal{G}_{i,n}) = \frac{(Y_{i,n}^H)^2}{p_{i,n}} + \frac{(Y_{i,n}^L)^2}{1-p_{i,n}} \leq \frac{2}{\eta}.$$

Thus

$$\mathbb{E}\left(\widehat{\tau}_{\mathrm{local},n} - \tau_n\right)^2 = \frac{\mathrm{Var}(\psi_{i,n})}{n} \leq \frac{2}{n\eta} \leq \frac{2}{n\eta^2}.$$

If $\mathrm{Var}(\psi_{i,n}) \geq v_0$, the same identity gives the stated v_0/n lower bound. Combining the two parts at $K_n \asymp n$ proves the separation. □

Proof of Proposition I.1. For fixed history value $\bar{h}$, the positive exact-arm propensities sum to one:

$$\sum_{a \in \mathcal{A}_n(\bar{h})} e_n(a \mid \bar{h}) = 1.$$

If every one of the $K_n(\bar{h})$ positive terms exceeded $1/K_n(\bar{h})$, their sum would exceed one. Hence the minimum is at most $1/K_n(\bar{h})$, proving (118). If the minimum were uniformly bounded below by $\eta > 0$, then $1 \geq K_n(\bar{h})\eta$, so $K_n(\bar{h}) \leq 1/\eta$. The final claim is the contrapositive. □

Proof of Theorem I.2. We give an explicit Assouad construction. In the conditional exact-arm subexperiment at history value $\bar{h}$, set $K = K_n(\bar{h})$. Let

$$\delta = \min\left\{M, \frac{\sigma}{4}\sqrt{\frac{K}{n}}\right\}$$

and index a finite submodel by $\omega = (\omega_1, \ldots, \omega_K) \in \{-1, 1\}^K$, with $\mu_a^\omega = \delta\omega_a$. Write P_ω for the law of the n observations under this parameter. For coordinate j, let $\omega^{(j)}$ denote ω with its jth sign reversed.

For one observation, P_ω and $P_{\omega^{(j)}}$ differ only when $A = j$, an event of probability $1/K$. Conditional on that event, the two normal means differ by 2δ. Therefore

$$\mathrm{KL}(P_\omega, P_{\omega^{(j)}}) = n\frac{1}{K}\frac{(2\delta)^2}{2\sigma^2} = \frac{2n\delta^2}{K\sigma^2}. \tag{114}$$

Pinsker's inequality and the choice of δ imply

$$\mathrm{TV}(P_\omega, P_{\omega^{(j)}}) \leq \sqrt{\frac{1}{2}\mathrm{KL}(P_\omega, P_{\omega^{(j)}})} = \sqrt{\frac{n\delta^2}{K\sigma^2}} \leq \frac{1}{4}.$$

Given any estimator $\widehat{\mu}$, define a sign estimator by $\widehat{\omega}_j = 1$ when $\widehat{\mu}_j \geq 0$ and -1 otherwise. Whenever $\widehat{\omega}_j \neq \omega_j$, $|\widehat{\mu}_j - \delta\omega_j| \geq \delta$. Hence

$$L_K(\widehat{\mu}, \mu^\omega) \geq \frac{\delta^2}{K}\sum_{j=1}^{K}\mathbf{1}\{\widehat{\omega}_j \neq \omega_j\}. \tag{115}$$

Average the right-hand side uniformly over the hypercube. For each j, pair every ω with $\omega^{(j)}$. The sum of the two testing errors for deciding the jth sign is at least $1-\mathrm{TV}(P_\omega, P_{\omega^{(j)}}) \geq 3/4$. Therefore

$$2^{-K}\sum_{\omega}\mathbb{E}_\omega\sum_{j=1}^{K}\mathbf{1}\{\widehat{\omega}_j \neq \omega_j\} \geq \frac{K}{2}\left(1-\frac{1}{4}\right) = \frac{3K}{8}.$$

Combining this inequality with (115), and using that the supremum dominates the hypercube average, gives

$$\sup_{\mu\in[-M,M]^K}\mathbb{E}_\mu L_K(\widehat{\mu},\mu) \geq \frac{3\delta^2}{8}.$$

Finally,

$$\delta^2 = \min\left\{M^2, \frac{\sigma^2 K}{16n}\right\} \geq \frac{1}{16}\min\left\{M^2, \frac{\sigma^2 K}{n}\right\}.$$

Taking the infimum over estimators proves (121) with $c_0 = 3/128$. □

Proof of Theorem I.3. All expectations in this proof are conditional on $E = 1$, and the ordered candidate pair is fixed before the final binary display draw. By consistency, $DY = DY(1)$ and $(1-D)Y = (1-D)Y(0)$. Conditional on $\mathcal{G} = \sigma\{Z, U, Y(0), Y(1)\}$, Assumption A.2 gives $\mathbb{E}(D \mid \mathcal{G}) = p(Z,U)$. Positivity permits division. Consequently,

$$\mathbb{E}\left\{\frac{DY}{p(Z,U)}\,\middle|\,\mathcal{G}\right\} = Y(1)\frac{\mathbb{E}(D\mid\mathcal{G})}{p(Z,U)} = Y(1),$$
$$\mathbb{E}\left\{\frac{(1-D)Y}{1-p(Z,U)}\,\middle|\,\mathcal{G}\right\} = Y(0)\frac{1-\mathbb{E}(D\mid\mathcal{G})}{1-p(Z,U)} = Y(0).$$

Taking conditional expectations given $E = 1$ and subtracting proves (122). The derivation conditions on the realized eligible candidate pair through Z and imposes no cross-unit recurrence or restriction on its generation law beyond the fixed-pair binary draw. □

H.3 Proof of the impossibility result

Proof of Theorem B.2. Consider one eligible ordered binary pair, so $E = 1$ throughout this construction, with local arm $D = 1$ denoting display of the h-higher candidate and $D = 0$ denoting display of the h-lower candidate. Let $U \sim \mathrm{Bernoulli}(1/2)$ and, conditional on U, let

$$\mathbb{P}(D = 1 \mid U = 0) = \frac{1}{4}, \qquad \mathbb{P}(D = 1 \mid U = 1) = \frac{3}{4}.$$

Set $Y(0) = 0$ almost surely. Current replay is independent of all historical variables and returns the h-higher candidate from this same ordered pair with probability $q = 1/2$.

In World 1, let $Y(1) = 1 - U$. In World 2, let $Y(1) = 0$ when $U = 0$ and, when $U = 1$, let $Y(1) \sim \mathrm{Bernoulli}(1/3)$ independently of the display draw conditional on U. In both worlds, conditional on U the display is independent of $(Y(0), Y(1))$, so Assumption A.2 holds with propensities $1/4$ and $3/4$. Consistency gives $Y = DY(1)$.

In World 1,

$$\mathbb{P}(D = 1, Y = 1) = \mathbb{P}(U = 0)\mathbb{P}(D = 1 \mid U = 0) = \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{8}.$$

In World 2,

$$\mathbb{P}(D = 1, Y = 1) = \mathbb{P}(U = 1)\mathbb{P}(D = 1 \mid U = 1)\mathbb{P}\{Y(1) = 1 \mid U = 1\} = \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{1}{3} = \frac{1}{8}.$$

Under either world, $\mathbb{P}(D = 1) = \frac{1}{2}(\frac{1}{4} + \frac{3}{4}) = 1/2$ and $Y = 0$ whenever $D = 0$. Hence both observable laws assign probabilities $1/8$, $3/8$, and $1/2$ to $(D, Y) = (1,1), (1,0), (0,0)$, respectively. The replay law is also identical and known exactly. In particular $\mathbb{E}(D \mid q) = 1/2 = q$. Since the odds of $1/4$ and $3/4$ are $1/3$ and 3, their ratio to the replay odds 1 is bounded by 3, which is equivalent to $\Gamma \geq \log 3$.

The eligible-pair effects are

$$\tau_h(P^{(1)}) = \mathbb{E}\{Y(1) - Y(0) \mid E = 1\} = \frac{1}{2},$$
$$\tau_h(P^{(2)}) = \mathbb{E}\{Y(1) - Y(0) \mid E = 1\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}.$$

Let T denote any estimator measurable with respect to the observable trace and any replay draws. Its distribution is identical in the two worlds. For every real t,

$$|t - 1/2| + |t - 1/6| \geq \frac{1}{3},$$

so, after expectation under their common observable law, at least one of the two absolute risks is at least $1/6$. Likewise,

$$(t - 1/2)^2 + (t - 1/6)^2 = 2(t - 1/3)^2 + \frac{1}{18} \geq \frac{1}{18},$$

so at least one squared risk is at least $1/36$. This proves every claim. □

Proof of Corollary I.4. Because the observable law is the same in the two worlds, the random interval C has the same distribution under both. Let $\mathcal{E}_j = \{\tau_h(P^{(j)}) \in C\}$, with $\tau_h(P^{(1)}) = 1/2$ and $\tau_h(P^{(2)}) = 1/6$. Under the common observable law, $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq 1 - \mathbb{P}(\mathcal{E}_1^c) - \mathbb{P}(\mathcal{E}_2^c) \geq 1 - 2\alpha$. On $\mathcal{E}_1 \cap \mathcal{E}_2$, the interval contains both points and therefore has length at least $1/3$. Hence $\mathbb{E}\,\mathrm{length}(C) \geq (1/3)\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq (1/3)(1 - 2\alpha)$. □

H.4 Validity, marginal-model sharpness, and duality

Proof of Proposition C.4. All expectations are under the eligible ordered-pair law conditional on $E = 1$, and both $p(Z, U)$ and $q(Z)$ refer to selection of the h-higher candidate from that same pair. We prove the result for $a = 1$; the control arm is identical after replacing p by $1 - p$. The drift bound implies $p_L(Z) \leq p(Z, U) \leq p_U(Z)$, hence $\ell_1(Z) \leq p(Z, U)^{-1} \leq u_1(Z)$. Conditional expectation preserves order, which gives (40).

Let f be bounded and measurable. By the tower property and the definition of $w_1^\star$,

$$\mathbb{E}\{Dw_1^\star(Y, Z)f(Z)\} = \mathbb{E}\left[D\frac{f(Z)}{p(Z, U)}\right].$$

Conditional on $(Z, U, Y(0), Y(1))$, Assumption A.2 gives $\mathbb{E}(D \mid Z, U, Y(0), Y(1)) = p(Z, U)$. Therefore the last display equals $\mathbb{E}f(Z)$, proving (41). Similarly, by consistency,

$$\mathbb{E}\{Dw_1^\star(Y, Z)Y\} = \mathbb{E}\left\{\frac{DY(1)}{p(Z, U)}\right\} = \mathbb{E}Y(1),$$

which proves (42) under the eligible-pair law. Taking f successively as the bounded coordinates of ϕ_K shows $w_a^\star \in \mathcal{W}_{a,K}$. Historical local positivity makes the projected weights essentially bounded; versions outside the observed arm support may be assigned the bounded lower envelope. The mean identity uses the explicitly assumed integrability of the potential outcomes. This population projection depends on the fixed replay anchor. □

Proof of Theorem C.6. Nonemptiness is assumed in the theorem; convexity follows from linearity of the moment constraints and convexity of the pointwise interval. The envelope bound in Assumption C.2 permits application of Lemma H.1 with $b_k = I_a\phi_k(Z)$ and $c = I_aY$. The requirement that w be $\sigma(Y, Z)$-measurable is also weak-$\star$ closed: this subspace is the annihilator of $\{h \in L^1(P) : \mathbb{E}(h \mid Y, Z) = 0\}$. Intersecting with it preserves compactness. The feasible set is weak-$\star$ compact and the objective is weak-$\star$ continuous, so both extrema are attained.

The image $M_a = \{\mathbb{E}(I_awY) : w \in \mathcal{W}_{a,K}\}$ is a compact convex subset of $\mathbb{R}$, hence a closed interval with endpoints $(\underline{\mu}_{a,K}, \overline{\mu}_{a,K})$. By Definition C.5, the arm weights are variation independent, so the image of $(w_1, w_0) \mapsto \mu_1(w_1) - \mu_0(w_0)$ is the Minkowski difference $M_1 - M_0 = [\underline{\mu}_{1,K} - \overline{\mu}_{0,K}, \overline{\mu}_{1,K} - \underline{\mu}_{0,K}]$. Every point in this interval is attained because M_1 and M_0 are intervals. Thus (48) is exactly the identified set for τ_h under the stated marginal model. □

Lemma H.3 (Pointwise minimization of the Lagrangian)**.** *For real r and $\ell \le u$,*

$$\inf_{w\in[\ell,u]} wr = \ell r_+ + u r_-, \qquad \sup_{w\in[\ell,u]} wr = u r_+ + \ell r_-.$$

Proof. If $r \ge 0$, multiplication by r preserves order and the minimum and maximum occur at ℓ and u; if $r < 0$, the order reverses. The displayed formulas combine the two cases. □

Proof of Theorem C.7. Fix an arm and abbreviate

$$\mathcal{C} = \{w \in L^\infty(P) : w \text{ is } \sigma(Y,Z)\text{-measurable},\ \ell \le w \le u\},$$
$$A(w) = \mathbb{E}(Iw\phi), \qquad J(w) = \mathbb{E}(IwY).$$

By Lemma H.1, $\mathcal{C}$ is weak-$\star$ compact and convex, while A and J are weak-$\star$ continuous. Hence the moment image $\mathcal{M} = A(\mathcal{C}) \subset \mathbb{R}^K$ is compact and convex.

We first verify that the target moment m is an interior point of $\mathcal{M}$. Let w° and δ_K be supplied by Assumption C.3, and put $G = \mathbb{E}\{I\phi(Z)\phi(Z)^\top\}$. For $b \in \mathbb{R}^K$, define

$$a_b = G^{-1}b, \qquad w_b = w^\circ + a_b^\top \phi.$$

Because $\|\phi\|_2 \le C_K$ and G is nonsingular, there is $r > 0$ such that $\|b\|_2 < r$ implies $|a_b^\top \phi| \le \delta_K$ almost surely. For such b, $w_b \in \mathcal{C}$ and

$$A(w_b) = A(w^\circ) + Ga_b = m + b.$$

Thus an open Euclidean ball around m is contained in $\mathcal{M}$.

Define the perturbation value function

$$V(t) = \min\{J(w) : w \in \mathcal{C},\ A(w) = t\}, \qquad t \in \mathcal{M}.$$

The minimum is attained by compactness. The function V is convex: if w_j minimizes at t_j, then the convex combination of w_1, w_2 is feasible at the corresponding convex combination of t_1, t_2. Since V is finite on a neighborhood of m, the subdifferential $\partial V(m)$ is nonempty; see, for example, the supporting-hyperplane theorem for finite convex functions (Rockafellar, 1970). Choose $\lambda^\star \in \partial V(m)$. For every $w \in \mathcal{C}$,

$$J(w) \ge V\{A(w)\} \ge V(m) + (\lambda^\star)^\top\{A(w) - m\}.$$

Consequently,

$$V(m) \le (\lambda^\star)^\top m + \mathbb{E}\big[Iw\{Y - (\lambda^\star)^\top\phi\}\big] \qquad \text{for every } w \in \mathcal{C}.$$

Taking the infimum over $w \in \mathcal{C}$ gives

$$V(m) \le (\lambda^\star)^\top m + \inf_{w\in\mathcal{C}} \mathbb{E}\big[Iw\{Y - (\lambda^\star)^\top\phi\}\big]. \tag{116}$$

The reverse inequality holds for every $\lambda \in \mathbb{R}^K$ by weak duality: if $A(w) = m$, then

$$\lambda^\top m + \mathbb{E}[Iw\{Y - \lambda^\top\phi\}] = J(w).$$

Therefore equality holds in (116), and $\lambda^\star$ is a dual maximizer.

It remains to evaluate the inner infimum. With $r(O) = Y - \lambda^\top\phi(Z)$, decomposability of the order interval and Lemma H.3 give

$$\inf_{w\in\mathcal{C}} \mathbb{E}(Iwr) = \mathbb{E}\{I[\ell r_+ + u r_-]\}.$$

Since $\lambda^\top m = \mathbb{E}\{\lambda^\top\phi(Z)\}$, this is exactly $\mathbb{E}g^-_{a,\lambda}(O)$ and proves (53). Applying the same argument to the lower problem with outcome $-Y$, followed by the change of variable $\lambda \mapsto -\lambda$, yields (54) and an attaining minimizer.

For later statistical use, the optimizing level sets are bounded. Indeed, for any λ let $\widetilde{w} = w^\circ - \delta_K \operatorname{sgn}(Y - \lambda^\top\phi)$ everywhere. This choice remains $\sigma(Y,Z)$-measurable. Then $\widetilde{w} \in \mathcal{C}$ and feasibility of w° gives

$$\mathbb{E}g^-_{a,\lambda} \le J(w^\circ) - \delta_K\mathbb{E}\{I|Y - \lambda^\top\phi|\}.$$

Writing $\lambda = tv$ with $\|v\|_2 = 1$, boundedness of Y and Lemma H.2 imply

$$\mathbb{E}\{I|Y - tv^\top \phi|\} \geq |t|\kappa_a - M.$$

Hence the lower dual objective tends to $-\infty$ as $\|\lambda\|_2 \to \infty$. The sign-reversed perturbation gives coercivity of the upper objective toward $+\infty$. Thus both optimizer sets are nonempty and compact, and a finite radius R_K can contain them.

Finally, for any feasible w and any λ, pointwise minimization and maximization yield $\mathbb{E}g^-_{a,\lambda} \leq J(w) \leq \mathbb{E}g^+_{a,\lambda}$. Taking the appropriate extrema over w proves (55). □

Proof of Corollary C.8. All statements here are under the eligible ordered-pair law. If $\Gamma = 0$, then $p_L = p_U = q$. Hence the treated box is the singleton $\{q^{-1}\}$ and the control box is the singleton $\{(1-q)^{-1}\}$. Under exact historical randomization, Proposition C.4 verifies the calibration moments, so these unique functions are feasible. Substitution into the arm means yields (57) for τ_h; the result does not depend on K. □

Proof of Theorem C.10. Nestedness of the basis implies $\mathcal{W}_{a,K+1} \subseteq \mathcal{W}_{a,K}$. Therefore lower endpoints are nondecreasing and upper endpoints are nonincreasing. Uniform boundedness of the envelopes puts all feasible sets in one weak-$\star$ compact ball.

Let $v_K = \underline{\mu}_{a,K}$ and $v = \lim_K v_K$. Choose a minimizer $w_K \in \mathcal{W}_{a,K}$. By weak-$\star$ compactness, a subnet converges to some w_∞. For each fixed J, all sufficiently large members of the subnet lie in the weak-$\star$ closed set $\mathcal{W}_{a,J}$, so $w_\infty \in \cap_J \mathcal{W}_{a,J}$. Define $r(Z) = \mathbb{E}\{I_a w_\infty(Y, Z) \mid Z\} - 1$. It is bounded. The moment restrictions give $\mathbb{E}\{r(Z)b(Z)\} = 0$ for every b in the union of the sieve spans. By density in $L^1(P_Z)$ and boundedness of r, the equality extends to every $b \in L^1(P_Z)$. Taking $b = \operatorname{sgn}(r)$ (or bounded approximations) gives $r = 0$ almost surely. Hence $w_\infty \in \mathcal{W}_{a,\infty}$.

Weak-$\star$ continuity of the objective gives along the convergent subnet $\mathbb{E}(I_a w_K Y) \to \mathbb{E}(I_a w_\infty Y)$. Thus $v \geq \underline{\mu}_{a,\infty}$ because w_∞ is fully feasible. Conversely, $\mathcal{W}_{a,\infty} \subseteq \mathcal{W}_{a,K}$ for every K, so $v_K \leq \underline{\mu}_{a,\infty}$ and therefore $v \leq \underline{\mu}_{a,\infty}$. Equality follows. The upper endpoint proof is identical with maxima.

An interval's Hausdorff distance is the maximum absolute difference of corresponding endpoints. Endpoint convergence for both arms therefore implies Hausdorff convergence of $\mathcal{I}_{\Gamma,K}$. □

Proof of Proposition D.1. The first statement is the conditional union bound. For the second, let $Z \sim \mathrm{Unif}[0,1]$ and $q(z) = 1/2$. Set the random band equal to $[1/2, 1/2]$ on $\{Z_1, \ldots, Z_n\}$ and $[3/5, 3/5]$ elsewhere. Its logged-context coverage is one, but its population coverage is zero because the realized finite set has P_Z-measure zero. The measurable band therefore gives a counterexample to the population implication. □

H.5 Unbiased replay and held-out certification

Proof of Proposition I.5. Conditional on (D_i, Z_i), $T_{i,r}$ has the negative-binomial stopping-time distribution for the number of Bernoulli trials required to observe r successes with probability $q_{D_i}(Z_i)$. It is the sum of r independent geometric waiting times on $\{1, 2, \ldots\}$, so

$$\mathbb{E}(T_{i,r} \mid D_i, Z_i) = \frac{r}{q_{D_i}(Z_i)}, \qquad \mathrm{Var}(T_{i,r} \mid D_i, Z_i) = \frac{r\{1 - q_{D_i}(Z_i)\}}{q_{D_i}(Z_i)^2}.$$

This proves (125). Conditional independence of replay and the historical outcome gives

$$\mathbb{E}\left[(2D_i - 1)Y_i \frac{T_{i,r}}{r} \,\middle|\, D_i, Y_i, Z_i\right] = (2D_i - 1)\frac{Y_i}{q_{D_i}(Z_i)}.$$

By the explicitly assumed consistency and historical conditional randomization, exact replay makes the conditional expectation of the right side the IPW representation in Corollary C.8, hence equal to τ_h. For independent eligible pairs, the conditional variance contributed solely by replay is

$$\frac{1}{n^2}\sum_{i=1}^{n} Y_i^2 \frac{1 - q_{D_i}(Z_i)}{r q_{D_i}(Z_i)^2} \leq \frac{M^2}{nr\eta_q^2}.$$

□

Proof of Proposition D.2. Conditional on the training sigma-field, $\widehat{\lambda}$ is fixed and the evaluation observations are i.i.d. from P. Linearity of conditional expectation gives both equalities. If $|g| \le B_K$, then $\mathrm{Var}(g \mid \mathcal{T}) \le \mathbb{E}(g^2 \mid \mathcal{T}) \le B_K^2$, and independence gives the variance bound for the average.

For optimism, for every fixed λ, $\sup_{\lambda'} P_m g^-_{\lambda'} \ge P_m g^-_{\lambda}$. Taking expectations and then the supremum over λ yields $\mathbb{E}\sup_{\lambda'} P_m g^-_{\lambda'} \ge \sup_{\lambda} P g^-_{\lambda}$. The infimum inequality follows after reversing all inequalities. □

Lemma H.4 (Uniform score bound)**.** *If* $0 \le Y \le M$, $\|\phi(Z)\|_2 \le C_K$, $\|\lambda\|_2 \le R_K$, *and* $u_a(Z) \le \bar{w}$, *then*

$$|g^{\pm}_{a,\lambda}(O)| \le B_K := R_K C_K + \bar{w}(M + R_K C_K).$$

Proof. Use $|\lambda^\top \phi| \le R_K C_K$, $|(Y - \lambda^\top \phi)_\pm| \le M + R_K C_K$, $I_a \le 1$, and $\ell_a \le u_a \le \bar{w}$. □

Proof of Theorem D.3. All population quantities refer to the eligible-pair law P. The independent replay sample generates a measurable band whose event $\mathcal{E}^{\mathrm{pop}}_{\mathrm{rep}}$ in (63) has failure probability at most α_{rep}. On this event, monotonicity of $q \mapsto \mathrm{expit}\{\mathrm{logit}(q) \pm \Gamma\}$ implies that the inflated historical envelope contains the true envelope for P_Z-almost every context. Thus every population-feasible weight remains feasible for the inflated model, and $\mathcal{I}_{\Gamma,K}(P,q) \subseteq \widetilde{\mathcal{I}}_{\Gamma,K}$.

Fix a fold ℓ, arm a, and direction. Condition on the independent replay sample and the historical observations outside I_ℓ. The band, the learned dual vector, and its score function are fixed, while the evaluation observations remain iid from P. The deterministic bound B_K applies for every realization of this training sigma-field. Empirical Bernstein therefore gives, with conditional probability at least $1 - \delta$,

$$P_{I_\ell} g^-_{a,\widehat{\lambda}^-_{a,-\ell}} - \rho_{m_\ell}(\widehat{v}^-_{a,\ell}, B_K, \delta) \le P g^-_{a,\widehat{\lambda}^-_{a,-\ell}},$$

and the upper-score mean plus its radius is at least its population expectation. Here P integrates a fresh observation with the replay band and learned dual held fixed. Integrating these conditional inequalities and applying a union bound over the $4L$ fold-arm-direction events gives failure probability at most $4L\delta$. No conditioning on the evaluation contexts or on the replay-coverage event is used.

On the intersection of these Bernstein events and $\mathcal{E}^{\mathrm{pop}}_{\mathrm{rep}}$, weak duality for the nonempty inflated model gives $P g^-_{a,\widehat{\lambda}} \le \widetilde{\underline{\mu}}_{a,K}$ and $P g^+_{a,\widehat{\lambda}} \ge \widetilde{\overline{\mu}}_{a,K}$. Convexly averaging with weights π_ℓ yields $\widehat{L}_a \le \widetilde{\underline{\mu}}_{a,K}$ and $\widehat{U}_a \ge \widetilde{\overline{\mu}}_{a,K}$, and hence the asserted interval inclusion. The total failure probability is at most $\alpha_{\mathrm{rep}} + 4L\delta = \alpha$. Under the clause-specific conditions of Proposition C.4, every compatible latent effect is covered by the population interval. The argument uses weak duality and therefore does not require exact optimization or strict interiority. □

Proof of Corollary D.4. At known-q zero drift, Corollary C.8 shows that the treated and control boxes collapse to the unique feasible weights q^{-1} and $(1-q)^{-1}$. Thus $\mathcal{I}_{0,K} = \{\tau_h\}$ for every K, even though strict interiority is impossible. By the explicit iid-history, fixed-fold, and fold-complement training construction, conditioning on each fold's historical training observations fixes the returned score while its held-out evaluation observations remain iid from P. The assumed $B_{K,0}$ bound permits the same four-direction empirical-Bernstein inequalities as in the proof of Theorem D.3, now with no replay failure event. Weak duality for the nonempty collapsed classes and a union bound over $4L$ events give failure probability at most $4L\delta = \alpha$. □

Proof of Corollary D.5. Condition on the sigma-field generated jointly by the pilot, final audit, and replay information and on a fold's historical training complement. By the explicit iid-history, fixed-fold construction and the stated joint independence, its evaluation observations remain iid from P, while the random radius, replay band, returned dual vector, and score function are fixed. The bound B_K^{aud} in (76) therefore permits the same conditional empirical-Bernstein argument as in Theorem D.3.

On $\mathcal{E}_{\mathrm{aud}}$ and the population replay-band event, monotonicity in both the replay endpoints and the drift radius makes the random inflated envelope contain the fixed population envelope generated

by $\Gamma^\star$ for P_Z-almost every context. Hence every weight in the fixed class induced by $(P, q, \Gamma^\star)$ is feasible for the random inflated class, so

$$\mathcal{I}_{\Gamma^\star,K}(P,q) \subseteq \widetilde{\mathcal{I}}_{\widehat{\Gamma}^U,K}.$$

Weak duality and the $4L$ Bernstein events give the second inclusion. A union bound requires no independence between the audit and replay events and yields total failure probability at most $\alpha_{\mathrm{aud}} + \alpha_{\mathrm{rep}} + 4L\delta = \alpha$. Finally, Proposition C.4 places every compatible latent effect obeying the fixed true bridge inside $\mathcal{I}_{\Gamma^\star,K}(P,q)$. □

H.6 Asymptotic normality

Lemma H.5 (Second-order effect of dual estimation)**.** *Under Assumptions C.2 and D.6, for every endpoint objective $G(\lambda) = Pg_\lambda$ and a root-n consistent optimizer $\widehat{\lambda}$,*

$$G(\widehat{\lambda}) - G(\lambda^\star) = O_p(n^{-1}), \qquad (P_m - P)(g_{\widehat{\lambda}} - g_{\lambda^\star}) = O_p(n^{-1})$$

when the evaluation sample of size $m \asymp n$ is independent of $\widehat{\lambda}$.

Proof. The first statement is Taylor's theorem around the interior optimum: $\nabla G(\lambda^\star) = 0$, the Hessian is continuous locally, and $\|\widehat{\lambda} - \lambda^\star\| = O_p(n^{-1/2})$.

For the second, positive and negative parts are one-Lipschitz. Hence, for a deterministic constant C depending only on C_K and $\bar{w}$, $|g_\lambda - g_{\lambda'}| \le C\|\lambda - \lambda'\|_2$ almost surely. Conditional on $\widehat{\lambda}$, the variance of the empirical-process difference is at most $C^2\|\widehat{\lambda} - \lambda^\star\|_2^2/m = O_p(n^{-2})$. Conditional Chebyshev therefore gives the result. □

Proof of Theorem D.7. Write $r_{nB} = n^{-1/2} + B^{-1/2}$ and consider one endpoint. Uniform nonsingularity and joint differentiability imply, by the implicit-function theorem, $\|\lambda^\star(\widehat{q}) - \lambda^\star(q)\| = O_p(B^{-1/2})$. All arguments below take place on the event $\widehat{q} \in \mathcal{Q}$, whose probability tends to one. Conditional on replay and a fold's historical training observations, expand its held-out score about $\lambda^\star(\widehat{q})$:

$$\begin{aligned} P_{I_\ell} g_{\widehat{\lambda}_{-\ell}}(\widehat{q}) - \mu(\widehat{q}) = {} & (P_{I_\ell} - P) g_{\lambda^\star(\widehat{q})}(\widehat{q}) \\ & + (P_{I_\ell} - P)\{g_{\widehat{\lambda}_{-\ell}}(\widehat{q}) - g_{\lambda^\star(\widehat{q})}(\widehat{q})\} \\ & + P\{g_{\widehat{\lambda}_{-\ell}}(\widehat{q}) - g_{\lambda^\star(\widehat{q})}(\widehat{q})\}. \end{aligned}$$

The last term is $O_p(n^{-1})$ by the stationary-point Taylor expansion, uniformly in the local replay argument. The preceding empirical-process term is $O_p(n^{-1})$ by the Lipschitz score bound, the assumed training rate, and conditional Chebyshev. Replacing the leading score by $g_{\lambda^\star(q)}(q)$ costs $O_p(n^{-1/2}B^{-1/2})$, again by conditional variance and local Lipschitz continuity. Averaging the fixed folds therefore gives

$$\widehat{\mu}(\widehat{q}) - \mu(\widehat{q}) = (P_n - P) g_{\lambda^\star(q)}(q) + O_p(n^{-1} + n^{-1/2}B^{-1/2}).$$

The envelope theorem and a second-order expansion yield

$$\mu(\widehat{q}) - \mu(q) = \sum_j d_j(\widehat{q}_j - q_j) + O_p(B^{-1}).$$

The derivative integrands displayed above follow directly by differentiating the weight bounds; the optimizer derivative vanishes because the population first-order condition is zero. The total remainder is $O_p(n^{-1} + B^{-1} + n^{-1/2}B^{-1/2}) = o_p(r_{nB})$. Substituting the independent Bernoulli sample means and replacing B_j/B by its positive limit gives (81). The historical and replay arrays are independent and bounded, so the multivariate central limit theorem gives covariance $V_O/n + V_R/B$ and the stated limit when $B/n \to \kappa \in (0, \infty)$. Held-out laws of large numbers, boundedness, and consistency of the fitted arguments prove consistency of the score and derivative covariance estimators. Linear transformation yields the two effect-endpoint influence functions. The resulting limit is a statement about convergence in distribution. □

Appendix remark (replay designs beyond finite strata). If a different replay estimator has a joint asymptotically linear expansion and the endpoint map is Hadamard differentiable at the corresponding pair-level replay law, its influence contribution can be obtained by the functional delta method. The main replay result above uses the explicit finite-stratum Bernoulli sample; the extension applies to other replay designs under these additional conditions.

Proof of Proposition D.8. Let $t = \mathbb{P}(D = 1, Y = 1) = 8/15$ and $s = \mathbb{P}(D = 1, Y = 0) = 1/15$. The treated box is $[\ell, u] = [3/2, 3]$. Historical $p = 3/5$ satisfies the bridge strictly, and the constant calibrated weight $5/3$ is strictly inside the box. If $x = tw_1$ is the treated mean, intercept calibration requires $sw_0 = 1 - x$, whence

$$L(t, s) = \max\{\ell t, 1 - us\}, \qquad U(t, s) = \min\{ut, 1 - \ell s\}.$$

Both lower branches equal $4/5$, while $U = 9/10$. The strict inequalities $\ell(t + s) = 9/10 < 1 < 9/5 = u(t + s)$ persist for empirical proportions with probability tending to one. The multinomial central limit theorem and continuity of the maximum give a limiting variable $\max\{X, Y\}$, where $X = (3/2)Z_t$, $Y = -3Z_s$, and (Z_t, Z_s) is the centered Gaussian limit of the two empirical proportions. Direct covariance calculation gives

$$\operatorname{Var} X = \operatorname{Var} Y = 14/25, \qquad \operatorname{Cov}(X, Y) = 4/25.$$

Thus $A = (X + Y)/2$ and $H = (X - Y)/2$ are independent centered Gaussian variables with variances $9/25$ and $1/5$. The maximum equals $A + |H|$; its strictly positive third centered moment proves non-Gaussianity. The endpoint directional derivative is $h \mapsto \max\{(3/2)h_t, -3h_s\}$, which is not linear. Equivalently, the population lower dual is flat for $\lambda \in [0, 1]$. This is a boundary of active-constraint selection despite strict feasibility. The statistic in this example is the empirical optimized value, not the cross-fitted score average; neither construction makes the population endpoint map differentiable at this law. The outer-certificate proof remains valid because it requires weak duality and concentration, not optimizer uniqueness. □

H.7 Minimax lower bounds

Lemma H.6 (Two-point quadratic lower bound)**.** *Let two experiments have laws $\mathbb{P}_0, \mathbb{P}_1$ and parameters θ_0, θ_1. Then every estimator T satisfies*

$$\max_{j=0,1} \mathbb{E}_j(T - \theta_j)^2 \geq \frac{(\theta_1 - \theta_0)^2}{16}\{1 - \mathrm{TV}(\mathbb{P}_0, \mathbb{P}_1)\}.$$

The same conclusion, up to the same universal constant, holds for squared Hausdorff loss when the two parameters are compact intervals separated in Hausdorff distance.

Proof. Let $\Delta = |\theta_1 - \theta_0|$ and define a test that selects the closer parameter to T. On a testing error, $|T - \theta_j| \geq \Delta/2$. The sum of the two testing errors is at least $1 - \mathrm{TV}(\mathbb{P}_0, \mathbb{P}_1)$. At least one risk is therefore at least $(\Delta/2)^2\{1 - \mathrm{TV}\}/2$, which is stronger than the displayed constant. For intervals, use a closest-parameter test under Hausdorff distance and the triangle inequality. □

Proof of Theorem D.9. The treated intercept constraint is $tw_1 + (1-t)w_0 = 2$ and the treated box is $[\ell(q), u(q)]$, with $u(q) = (2-q)/q$. At $(t, q) = (1/2, 1/2)$, fixing either outcome weight at $\ell = 3/2$ leaves the other at $5/2$, strictly inside $[3/2, 3]$. These strict inequalities persist on $\mathcal{C}_\epsilon$ (and hold for $\epsilon \leq 1/100$). The lower endpoint fixes $w_1 = \ell(q)$ and the upper endpoint fixes $w_0 = \ell(q)$, giving the displayed formulas. Control outcomes are zero. Historical $p = 1/2$ is inside the log-odds bridge throughout this neighborhood. Outcome values are 0 and M, so all bounds hold for every $M > 0$.

Let $N = \sum_i D_i$, set $\widetilde{t} = \sum_i D_i Y_i/(MN)$ when $N > 0$ and $\widetilde{t} = 1/2$ when $N = 0$, and let $\widetilde{q}$ be the replay sample mean. Project each coordinate onto $[1/2 - \epsilon, 1/2 + \epsilon]$ to obtain $\widehat{t}, \widehat{q}$. Conditional on $N > 0$, $\widetilde{t}$ is a Bernoulli sample mean and its conditional MSE is at most $1/(4N)$. Since $N \sim \mathrm{Binomial}(n, 1/2)$,

$$E\{N^{-1}1(N > 0)\} \leq 2E\{(N + 1)^{-1}\} \leq \frac{4}{n + 1}, \qquad P(N = 0) = 2^{-n}.$$

Projection cannot increase distance from the true parameter. Consequently $E(\widehat{t}-t)^2 \le C/n$ and $E(\widehat{q}-q)^2 \le 1/(4B)$. The two endpoint maps have uniformly bounded derivatives on the compact rectangle. Substitution therefore gives $\sup E d_H^2\{\mathcal{I}(\widehat{t},\widehat{q}),\mathcal{I}(t,q)\} \le C(n^{-1}+B^{-1})$.

For the historical lower bound, fix $q=1/2$ and take $t_\pm = 1/2 \pm d/\sqrt{n}$, with $d>0$ small enough that both lie in the rectangle. Both endpoints have derivative $M\ell(1/2)/2 = 3M/4$ in t. The historical KL divergence is $n/2$ times the Bernoulli KL divergence, bounded by a fixed small constant for sufficiently small d; replay laws coincide. The two-point testing bound yields c/n. For the replay lower bound, fix $t=1/2$ and take $q_\pm = 1/2 \pm d/\sqrt{B}$. Endpoint derivatives in q are $-Mt/(4q^2)$ and $M(1-t)/(4q^2)$, nonzero and bounded away from zero on the rectangle. The historical laws coincide and the replay KL is bounded, so the same testing bound gives c/B. The union is contained in the same positive-radius class and establishes the set-risk lower bound.

For the point-decision sampling lower bounds choose the lower endpoint $\theta(t,q) = Mt\ell(q)/2$ in each testing subexperiment; it has the same nonzero derivatives. For the ambiguity lower bound, at each fixed observable law average the squared losses at the two endpoints to obtain at least $r(q)^2 \ge r_{\min}^2$. Taking the maximum of these valid restrictions proves the displayed lower bound. For the upper bound, use the midpoint of $\mathcal{I}(\widehat{t},\widehat{q})$. Its MSE for the true midpoint is $O(n^{-1}+B^{-1})$, and for every $\theta \in \mathcal{I}(t,q)$, $(T-\theta)^2 \le 2(T-\operatorname{mid}\mathcal{I})^2 + 2r(q)^2$. This proves the upper bound. □

H.8 Optimal bridge design

Proof of Proposition E.1. For both arms, differentiation of the logistic weight envelopes gives $\partial_{\Gamma_j}\ell_a = -\mathbf{1}\{Z \in C_j\}(\ell_a - 1)$ and $\partial_{\Gamma_j}u_a = \mathbf{1}\{Z \in C_j\}(u_a-1)$. Thus the lower score derivative is $I_a\mathbf{1}\{Z\in C_j\}\{-(\ell_a-1)(r_a^-)_+ + (u_a-1)(r_a^-)_-\}$, and the upper score derivative is $I_a\mathbf{1}\{Z\in C_j\}\{(u_a-1)(r_a^+)_+ - (\ell_a-1)(r_a^+)_-\}$. On the common compact set the residuals, envelopes, and their derivatives are uniformly bounded. Dominated convergence therefore differentiates the expected objectives, with a derivative continuous jointly in the radius and dual vector. For radii tending to the reference, compactness and continuity imply that every cluster point of a maximizing/minimizing dual vector optimizes the reference objective. Uniqueness makes these optimizers converge to the corresponding reference optimizer. Bounding the perturbed optimum using its own optimizer and the reference optimizer gives the ordinary envelope expansion, uniformly over unit admissible directions. Since $W = \sum_a(\mu_a^+ - \mu_a^-)$, subtracting lower derivatives from upper derivatives yields the formula. Every summand is nonnegative because $r_+ \ge 0$, $r_- \le 0$, and $u_a \ge \ell_a \ge 1$. This proof differentiates the envelopes directly; no no-atom residual assumption is required for this particular derivative. □

Proof of Theorem E.2. Condition on the independent pilot sigma-field, so all final counts are fixed. The stipulated freshness and independence make each final sample mean an ordinary mean of independent Bernoulli draws with the displayed cell probability. Hoeffding's inequality gives, for each historical or replay sample mean,

$$\mathbb{P}\{|\widehat{p}_j - p_j| > e_{p,j}\} \le \frac{\alpha_{\mathrm{aud}}}{2J}, \qquad \mathbb{P}\{|\widehat{q}_j - q_j| > e_{q,j}\} \le \frac{\alpha_{\mathrm{aud}}}{2J}.$$

A union bound over j and the two experiments proves simultaneous coverage of the clipped intervals in (94); clipping cannot remove a probability in $[\eta, 1-\eta]$. On this event, the true pair (p_j, q_j) is in the rectangle over which (95) maximizes, proving (96).

On $[\eta_0, 1-\eta_0]$, the derivative of logit is at most L_{η_0}. The absolute-difference map is one-Lipschitz in the ℓ_1 norm, so enlarging both coordinate intervals gives

$$0 \le \widehat{\Gamma}_j^U - \Gamma_j \le 2L_{\eta_0}(e_{p,j} + e_{q,j}),$$

which is (97). Substituting $B_j = \beta_j m_j$ yields (100). For fixed J and fixed confidence level, the sample-mean errors are $O_p(m_j^{-1/2})$ and $O_p(B_j^{-1/2})$. Clipping is nonexpansive and the logit is L_{η_0}-Lipschitz on the clipped interval, so the absolute radius error is bounded by a constant times the two sample-mean errors plus $e_{p,j} + e_{q,j}$. This proves the fixed-confidence root-budget order without requiring the coverage probability to tend to one. If the confidence level varies, the deterministic widths retain $\sqrt{\log(4J/\alpha_{\mathrm{aud}})}$; the same Lipschitz argument gives (99) with that factor. Under (98) this error tends to zero. Cell homogeneity converts the certified difference between cell

probabilities into a pointwise historical-to-replay bridge throughout each cell. The conditional probability bound is uniform in the pilot realization, so integrating over the pilot proves the unconditional statement. Joint independence from reserved history permits the resulting radius certificate to enter Corollary D.5. □

Proof of Theorem E.3. Here the cells partition the eligible ordered-pair population, and the objective is the first-order excess width of the identified set for τ_h. It is strictly convex on $(0,\infty)^J$ and strictly decreasing in every coordinate, so the unique minimizer uses the full budget. The Lagrangian is

$$\mathcal{L}(m,\lambda) = \sum_{j=1}^{J} a_j m_j^{-1/2} + \lambda \left(\sum_{j=1}^{J} \kappa_j m_j - C_{\mathrm{aud}} \right), \qquad \lambda > 0.$$

The first-order condition is $-(1/2)a_j m_j^{-3/2} + \lambda\kappa_j = 0$, whence

$$m_j = (a_j/(2\lambda\kappa_j))^{2/3}.$$

Imposing the budget gives $(2\lambda)^{-2/3}\sum_j a_j^{2/3}\kappa_j^{1/3} = C_{\mathrm{aud}}$, and substitution yields (104). Let $S = \sum_j a_j^{2/3}\kappa_j^{1/3}$. Then

$$\sum_j a_j (m_j^\star)^{-1/2} = \sum_j a_j \left\{ \frac{S}{C_{\mathrm{aud}}(a_j/\kappa_j)^{2/3}} \right\}^{1/2} = \frac{S^{3/2}}{\sqrt{C_{\mathrm{aud}}}},$$

which is (105). With lower bounds, complementary slackness fixes the cells at their lower bounds and the same calculation applies on the remaining active coordinates using the residual budget. □

Proof of Corollary E.4. Throughout this proof, $W(\mathbf{\Gamma}) = \mathrm{length}\{\mathcal{I}_{\mathbf{\Gamma},K}\}$ is the width of the τ_h identified set under the eligible ordered-pair law. Write $m_j = C_{\mathrm{aud}} x_j$ and define the positive budget simplex

$$\mathcal{X} = \left\{ x \in (0,\infty)^J : \sum_j \kappa_j x_j \le 1 \right\}.$$

For fixed $x \in \mathcal{X}$,

$$\mathbf{\Gamma}^{\mathrm{cmp}}(C_{\mathrm{aud}}x) - \mathbf{\Gamma}_\infty = C_{\mathrm{aud}}^{-1/2}(c_1 x_1^{-1/2}, \ldots, c_J x_J^{-1/2})^\top.$$

The Fréchet expansion (106) therefore gives, uniformly on every subset of $\mathcal{X}$ bounded away from its coordinate boundary,

$$\sqrt{C_{\mathrm{aud}}}\,[W\{\mathbf{\Gamma}^{\mathrm{cmp}}(C_{\mathrm{aud}}x)\} - W(\mathbf{\Gamma}_\infty)] = F(x) + o(1), \qquad F(x) = \sum_{j=1}^{J} s_j c_j x_j^{-1/2}. \tag{117}$$

The function F is continuous and strictly convex, uses the full budget at its minimum, and diverges whenever any coordinate tends to zero. Hence it has a unique minimizer $x^\star$ in the interior of the positive budget simplex. By Theorem E.3, $x^\star = m^\star/C_{\mathrm{aud}}$ and

$$F(x^\star) = \left\{ \sum_j (s_j c_j)^{2/3} \kappa_j^{1/3} \right\}^{3/2}.$$

We next show that a sequence whose exact widths approach the infimum to $o(C_{\mathrm{aud}}^{-1/2})$ cannot escape to the coordinate boundary; this does not assume attainment on the open allocation domain. The width W is coordinatewise nondecreasing because enlarging any drift radius enlarges both armwise feasible classes, weakly decreases lower endpoints, and weakly increases upper endpoints. If some m_j stays bounded along a subsequence, then the jth radius remains separated from $\Gamma_{j,\infty}$. Since $s_j > 0$, (106) implies that increasing only the jth radius by a sufficiently small fixed amount raises the width by a positive amount. Monotonicity then keeps the subsequential excess width bounded

away from zero, whereas any interior proportional allocation has excess width $O(C_{\mathrm{aud}}^{-1/2})$; such a subsequence cannot be optimal.

Once every $m_k \to \infty$, the radius increment h tends to zero. Positivity gives $s^\top h \geq (\min_k s_k)\|h\|_2$ for $h \geq 0$, so the remainder is $o(s^\top h)$. If some $x_j = m_j/C_{\mathrm{aud}} \to 0$, then $\sqrt{C_{\mathrm{aud}}}\, s_j c_j m_j^{-1/2} = s_j c_j x_j^{-1/2} \to \infty$. Hence no $O(C_{\mathrm{aud}}^{-1/2})$-optimal sequence can have $x_j \to 0$. Every such near-minimizing sequence therefore eventually lies in a compact subset of $\mathcal{X}$ bounded away from the coordinate boundary. Uniformity in (117) and a direct argmin comparison give

$$\inf_x \sqrt{C_{\mathrm{aud}}}\,[W\{\mathbf{\Gamma}^{\mathrm{cmp}}(C_{\mathrm{aud}}x)\} - W(\mathbf{\Gamma}_\infty)] = F(x^\star) + o(1),$$

which proves (107); evaluating at $x^\star$ proves (108).

Finally, coordinatewise rounding down preserves the budget and changes each positive, order-C_{aud} allocation by at most one. Its radius increment and the unrounded increment both have norm $O(C_{\mathrm{aud}}^{-1/2})$, and their difference is $O(C_{\mathrm{aud}}^{-3/2})$. Apply (106) separately at these two increments: linear terms differ by $O(C_{\mathrm{aud}}^{-3/2})$ and each remainder is $o(C_{\mathrm{aud}}^{-1/2})$. Thus their widths differ by $o(C_{\mathrm{aud}}^{-1/2})$, without assuming a local Lipschitz property between arbitrary nearby radii. □

Proof of Theorem E.5. Because $a_j = s_j c_j > 0$ and the pilot estimates are consistent, the event $\{\widehat{a}_j > 0 \text{ for all } j\}$ has probability tending to one, and $\widehat{a}_j \xrightarrow{p} a_j$. The normalized allocation map

$$T_j(a_1, \ldots, a_J) = \frac{(a_j/\kappa_j)^{2/3}}{\sum_{k=1}^J a_k^{2/3}\kappa_k^{1/3}}$$

is continuous on the positive orthant. Hence $\widehat{m}_j^{\mathrm{plug}}/C_{\mathrm{aud}} = T_j(\widehat{a}) \xrightarrow{p} T_j(a) = m_j^\star/C_{\mathrm{aud}}$, and the normalized plug-in allocation lies in a compact subset of the positive budget simplex with probability tending to one.

Let $F(x) = \sum_j s_j c_j x_j^{-1/2}$. The uniform expansion in the proof of Corollary E.4, applied on that compact subset, gives

$$C_{\mathrm{aud}}^{1/2}\left[W\{\mathbf{\Gamma}^{\mathrm{cmp}}(\widehat{m}^{\mathrm{plug}})\} - W(\mathbf{\Gamma}_\infty)\right] = F(\widehat{m}^{\mathrm{plug}}/C_{\mathrm{aud}}) + o_p(1) = F(m^\star/C_{\mathrm{aud}}) + o_p(1).$$

The same expansion at the oracle allocation gives

$$C_{\mathrm{aud}}^{1/2}\,[W\{\mathbf{\Gamma}^{\mathrm{cmp}}(m^\star)\} - W(\mathbf{\Gamma}_\infty)] = F(m^\star/C_{\mathrm{aud}}) + o(1).$$

Since $F(m^\star/C_{\mathrm{aud}}) > 0$, division proves (111). Integer rounding perturbs each allocation by $O(1)$ and, because every oracle cell receives order-C_{aud} audits, changes the width by $o(C_{\mathrm{aud}}^{-1/2})$. If lower bounds are imposed and the active sets agree, the same argument is applied to the free coordinates after subtracting the fixed lower-bound cost; the residual-budget allocation map is continuous on the corresponding positive orthant. □

I Supporting Results and Bridge Extensions

I.1 Supporting exact-arm and local statements

I.1.1 Boundary I: exact-arm overlap and replication

Proposition I.1 (Uniform exact-arm overlap is incompatible with arm-space expansion)**.** *Fix a history value $\bar{h}$ and suppose the conditional population/design support contains $K_n(\bar{h})$ exact treatments with positive global propensities $e_n(a \mid \bar{h})$. Then*

$$\min_{a \in \mathcal{A}_n(\bar{h})} e_n(a \mid \bar{h}) \leq \frac{1}{K_n(\bar{h})}. \tag{118}$$

Consequently, a fixed exact-arm overlap condition $\inf_n \inf_{\bar{h}} \min_{a \in \mathcal{A}_n(\bar{h})} e_n(a \mid \bar{h}) \geq \eta > 0$ forces $K_n(\bar{h}) \leq 1/\eta$ for every n and $\bar{h}$. If $K_n(\bar{h}) \to \infty$ for some sequence of history values, the global exact-treatment positivity constant must converge to zero.

The proposition divides unit probability mass among $K_n(\bar{h})$ exact identities. Semantic smoothness or a coarser representation would share information across them. The next theorem studies recovery of the unstructured exact-arm response surface even with known assignment probabilities.

Theorem I.2 (Known propensities do not restore exact-arm response recovery)**.** *For a fixed history value $\bar{h}$, let $K = K_n(\bar{h}) \geq 2$ and consider the conditional exact-arm subexperiment*

$$A_i \sim \mathrm{Unif}\{1, \ldots, K\}, \qquad Y_i \mid A_i = a \sim N(\mu_a, \sigma^2), \qquad \mu \in [-M, M]^K, \tag{119}$$

where the n observations are iid and the propensity $1/K$ and $\sigma > 0$ are known. For the average exact-arm response loss

$$L_K(\widehat{\mu}, \mu) = \frac{1}{K} \sum_{a=1}^{K} (\widehat{\mu}_a - \mu_a)^2, \tag{120}$$

there is a universal constant $c_0 > 0$ such that

$$\inf_{\widehat{\mu}} \sup_{\mu \in [-M,M]^K} \mathbb{E}_\mu L_K(\widehat{\mu}, \mu) \geq c_0 \min\left\{ M^2, \frac{\sigma^2 K_n(\bar{h})}{n} \right\}. \tag{121}$$

In particular, if $K_n(\bar{h})/n$ is bounded away from zero and M, σ are fixed, no estimator consistently recovers the exact-arm response surface, even under randomized treatment with perfectly known propensities.

Interpretation and boundary. The lower bound concerns recovery of a separate response for every exact treatment. A scalar average may remain estimable when $K_n(\bar{h})$ is large. Semantic representations change the parameter class by sharing information, while local randomization changes the estimand and the unit of design information. These routes address different aspects of the replication problem.

I.1.2 Identification from local candidate randomization

Theorem I.3 (Local randomization identifies a generated-treatment contrast)**.** *Suppose Assumptions A.1 and A.2 hold on the eligible population $E = 1$, the main setup's ordered binary pair is fixed before the display draw, the historical pair-level probability $p(Z, U)$ is observed, and $\mathbb{E}|Y(a)| < \infty$ for both arms. Then*

$$\tau_h = \mathbb{E}\left[\frac{DY}{p(Z,U)} - \frac{(1-D)Y}{1-p(Z,U)} \;\middle|\; E = 1 \right]. \tag{122}$$

The identity holds without any restriction on how the eligible pair is generated and without recurrence of either exact candidate across units, provided the final within-pair draw is randomized conditional on the complete production state.

Pair formation may depend on latent state associated with both potential outcomes. What identifies the local contrast is the randomized final display, conditional on the pair and complete production state. Averaging its Horvitz–Thompson scores recovers the algorithm-relative eligible-population effect. This distinction between pair generation and display randomization is the design-based foundation of logged counterfactual exposures.

Historical probability requirement. Equation (122) uses the pair-level probability that actually selected the historical candidate. Using a score reconstructed from a different model version requires a bridge to the historical selection law. The second boundary shows that arbitrarily accurate current replay cannot, by itself, repair this temporal mismatch.

Corollary I.4 (Minimum length of an honest interval)**.** *Let C be a random interval based on the observable trace and replay draws. If $\inf_{j=1,2} \mathbb{P}_{P^{(j)}}\{\tau_h(P^{(j)}) \in C\} \geq 1 - \alpha$ for $\alpha < 1/2$, then*

$$\mathbb{E}\{\mathrm{length}(C)\} \geq \frac{1}{3}(1 - 2\alpha), \tag{123}$$

where the expectation is the same under both observationally equivalent laws.

I.2 An exact-replay unbiased benchmark

The exact-replay case admits a finite-replay estimator that is useful as a benchmark. Conditional on an eligible logged interaction, repeatedly rerun the current mechanism on the same ordered pair until the historically displayed side has appeared $r \geq 1$ times. Let $T_{i,r}$ be the total number of replay draws required, using success probability $q_{D_i}(Z_i) = D_i q(Z_i) + (1 - D_i)\{1 - q(Z_i)\}$.

Proposition I.5 (Inverse-binomial unbiasedness)**.** *Assume consistency and historical conditional randomization for the eligible ordered pair, with $p(Z, U) = q(Z)$ almost surely, $q \in [\eta_q, 1 - \eta_q]$, and $0 \leq Y(a) \leq M$. Conditional on all historical observations, assume each replay stream consists of independent Bernoulli draws, the streams are mutually independent across units, and they are independent of historical treatment and outcome draws given their prescribed inputs. Then*

$$\widehat{\tau}_{\mathrm{IB}} = \frac{1}{n} \sum_{i=1}^{n} (2D_i - 1) Y_i \frac{T_{i,r}}{r} \tag{124}$$

is exactly unbiased for τ_h. Moreover,

$$\mathrm{Var}\left(\frac{T_{i,r}}{r} \,\middle|\, D_i, Z_i \right) = \frac{1 - q_{D_i}(Z_i)}{r q_{D_i}(Z_i)^2} \leq \frac{1}{r \eta_q^2}. \tag{125}$$

Consequently, conditional replay noise contributes at most $M^2/(nr\eta_q^2)$ to the variance of (124)*.*

This unbiasedness disappears when current replay differs from historical selection: $T_{i,r}/r$ remains unbiased for $1/q_{D_i}$, not for the required $1/p_{D_i}$. The appropriate nonzero-drift target is therefore the marginal trace identified set for τ_h rather than a corrected point estimator; every compatible latent historical effect is covered by this set through (50).

I.3 From semantic generator drift to propensity envelopes

An odds radius is natural when archived propensity pairs are available. For generative systems, a semantic metric may bound change in the historical output distribution. We develop this bridge for a fixed eligible ordered pair. Let V be a generated candidate in a compact metric space $\mathcal{V}$, let Q_z be the current replay distribution conditional on context z and the ordered pair, and let $c : \mathcal{V} \times \mathcal{V} \to [0, \infty]$ be a lower-semicontinuous transport cost. Suppose the unknown historical generator P_z lies in

$$\mathcal{P}_z(\rho_z) = \{P : W_c(P, Q_z) \leq \rho_z\}, \tag{126}$$

where W_c is the optimal-transport cost. For an event $B_z \subseteq \mathcal{V}$ corresponding to the h-higher candidate within the eligible ordered pair, define

$$p_U(z) = \sup_{P \in \mathcal{P}_z(\rho_z)} P(B_z), \qquad p_L(z) = \inf_{P \in \mathcal{P}_z(\rho_z)} P(B_z). \tag{127}$$

Theorem I.6 (Wasserstein-to-propensity dual envelope)**.** *Let $\mathcal{V}$ be compact metric, let $c : \mathcal{V}^2 \to [0, \infty)$ be finite and continuous with $c(v, v) = 0$, let Q_z be a probability law, and let $\rho_z \geq 0$. Suppose the selection event B_z is both open and closed in $\mathcal{V}$ (in particular, any event on a finite discrete space qualifies). Then*

$$p_U(z) = \inf_{\lambda \geq 0} \left[\lambda \rho_z + \mathbb{E}_{V \sim Q_z} \sup_{v \in \mathcal{V}} \{ \mathbf{1}(v \in B_z) - \lambda c(v, V) \} \right], \tag{128}$$

$$p_L(z) = 1 - \inf_{\lambda \geq 0} \left[\lambda \rho_z + \mathbb{E}_{V \sim Q_z} \sup_{v \in \mathcal{V}} \{ \mathbf{1}(v \notin B_z) - \lambda c(v, V) \} \right]. \tag{129}$$

Assume the resulting bounds are measurable in z and satisfy $0 < \eta_* \leq p_L(z) \leq p_U(z) \leq 1 - \eta_*$ almost surely. Define $\mathcal{I}_{[p_L, p_U], K}$ by the primal formulas (32)–(48) *using these direct probability bounds. If the corresponding marginal weight classes are nonempty, this interval is sharp for that marginal model. For latent-effect validity, additionally require, for almost every historical state (z, u), that its conditional generator $P_{z,u}$ lies in $\mathcal{P}_z(\rho_z)$ and satisfies $p(z, u) = P_{z,u}(B_z)$. Under consistency, historical conditional randomization and bounded outcomes, every compatible latent effect belongs to $\mathcal{I}_{[p_L, p_U], K}$. A bound only on a generator averaged over U does not imply this pointwise condition.*

The event-continuity condition covers finite candidate-output spaces and separated selection regions with continuous costs. Extending the result to arbitrary Borel selection events requires further conditions. The direct probability envelope can be asymmetric in log odds around q, so its interval is indexed by $[p_L, p_U]$. Historical conditional randomization remains an independent assumption.

Proof of Theorem I.6. Fix z and abbreviate $Q = Q_z$, $\rho = \rho_z$, and $f = \mathbf{1}_{B_z}$. Clopenness makes f continuous. The set Π_Q of probability couplings on $\mathcal{V}^2$ whose second marginal is Q is compact and convex in the weak topology. Both $C(\pi) = \int c(v, V)\, d\pi$ and $F(\pi) = \int f(v)\, d\pi$ are continuous. The identity coupling has zero cost, and the feasible set is compact. Minimizing transport cost for fixed marginals is attained, so optimizing over $\{\pi \in \Pi_Q : C(\pi) \le t\}$ gives the same value as the distributional transport ball. Denote that value by $v(t)$ for $t \ge 0$.

The function v is finite, nondecreasing and concave. For $\rho > 0$ it has a finite supergradient $\lambda \ge 0$ at ρ, since it is a finite concave function on $(0, \infty)$. Its support inequality implies $F(\pi) \le v(\rho) + \lambda\{C(\pi) - \rho\}$ for all couplings with positive cost, and by right limits also for zero-cost couplings. Hence

$$\lambda\rho + \sup_{\pi \in \Pi_Q} \{F(\pi) - \lambda C(\pi)\} \le v(\rho).$$

Weak duality gives the opposite inequality. For $\rho = 0$, compactness and continuity show $v(t) \downarrow v(0)$ as $t \downarrow 0$: any cluster point of maximizing couplings is feasible at zero and attains the limiting payoff. The dual value at zero is at most the dual value at any $t > 0$, hence at most $\lim_{t \downarrow 0} v(t) = v(0)$; weak duality again gives equality. At zero, equality of the dual value can hold without a finite minimizing multiplier.

For fixed λ, the continuous function $f(v) - \lambda c(v, V)$ attains its maximum on compact $\mathcal{V}$; a measurable maximizing selection as a function of V realizes

$$\sup_{\pi \in \Pi_Q} \{F(\pi) - \lambda C(\pi)\} = \mathbb{E}_{V \sim Q} \sup_{v \in \mathcal{V}} \{f(v) - \lambda c(v, V)\}.$$

This proves (128). The complement of a clopen event is also clopen; applying the same identity to $1 - f$ and using $\inf_P P(B_z) = 1 - \sup_P P(B_z^c)$ proves (129).

Uniform positivity bounds the induced inverse-weight boxes. Nonempty calibrated boxes have attained interval-valued objective images by the compactness argument in Theorem C.6, which does not need strict interiority once nonemptiness is given. Their variation-independent difference is exactly $\mathcal{I}_{[p_L, p_U], K}$. Under the additional state-conditional generator premise, $p_L(z) \le p(z, u) \le p_U(z)$ for almost every (z, u). The tower-conditioning proof of Proposition C.4 then puts each compatible latent effect in this interval. □

I.4 Auxiliary finite-action OPE extension

For a finite action set, let A be the logged action, Z the logged context, and $\pi(a \mid z)$ a fixed target policy. Write $p(a \mid z)$ for the historical logging propensity and $q(a \mid z) > 0$ for the nominal replay score. Assume $p(a \mid z) > 0$ wherever $\pi(a \mid z) > 0$. A multiplicative score envelope $|\log\{p(a \mid z)/q(a \mid z)\}| \le \Gamma(z)$ bounds the historical importance weight by

$$\underbrace{\frac{\pi(A \mid Z)}{q(A \mid Z)} e^{-\Gamma(Z)}}_{\ell_\pi(A,Z)} \le \frac{\pi(A \mid Z)}{p(A \mid Z)} \le \underbrace{\frac{\pi(A \mid Z)}{q(A \mid Z)} e^{\Gamma(Z)}}_{u_\pi(A,Z)}. \tag{130}$$

This follows by exponentiating the log-ratio bound and multiplying its reciprocal by $\pi(A \mid Z)/q(A \mid Z)$. The induced box uses likelihood ratios for individual actions; it is distinct from the binary-pair odds envelope in (28). For the injected-score diagnostic below, q need not define a normalized policy over all actions.

The auxiliary empirical calculation minimizes and maximizes $n^{-1} \sum_i w_i Y_i$ over this box, with stratum normalization $\sum_{i: Z_i \in S_j} w_i = n_j$, where S_j is a context stratum and n_j its sample size. Appendix J.8 records the implemented feasibility adjustment and sampling expansion; Section F.4 reports the real-log results. The calculation targets finite-action policy value. The main contrast τ_h additionally requires a pre-display generated pair and its historical selection law.

J Experimental Protocol, Additional Diagnostics, and Reproducibility

The semi-synthetic study samples n iid simulated historical traces with replacement from a fixed eligible confirmatory corpus; n counts traces from existing pairs and ratings. Within each trace, $D_i = 1$ denotes display of the h-higher candidate and $D_i = 0$ the h-lower candidate; historical $p(Z_i, U_i)$ and replay $q(Z_i)$ refer to that pair. Supplementary local diagnostics have their own eligible-pair populations and sample definitions. Exact-arm experiments use conditional support notation $\mathcal{A}_n(h)$ and $K_n(\bar{h})$. The Open Bandit Dataset provides an auxiliary fixed-action diagnostic because it lacks pre-display binary candidate pairs for the main τ_h estimand.

Throughout this appendix, "prespecified," "frozen," and "no selective rerun" refer to the recorded project provenance. The protocol or historical freeze manifest predates the completion manifest and retained outputs, and adverse outcomes are preserved where described.

J.1 Analytic solver for the stratified empirical program

For experiments using a saturated stratum basis, the empirical program has the following analytic solver. The lower-dimensional moment bases in Experiments 5–7 instead use their specified optimization routines. In a stratum s containing n_s observations from the relevant analysis population (the effective eligible-pair log for local experiments), the arm-a normalization constraint is

$$\sum_{i: Z_i \in s, D_i = a} w_i = n_s.$$

Let the selected observations have outcomes y_i, lower bounds ℓ_i, and capacities $c_i = u_i - \ell_i$. Starting from $w_i = \ell_i$, define residual mass $R_s = n_s - \sum_i \ell_i$. Feasibility is equivalent to $0 \le R_s \le \sum_i c_i$. The lower endpoint allocates residual mass greedily to observations in increasing order of y_i; the upper endpoint uses decreasing order. This is exact because the stratum problem is a continuous knapsack linear program, whose extreme point has at most one non-boundary weight.

Algorithm 3 Exact stratum endpoint by continuous knapsack

Require: Outcomes y_i, bounds $[\ell_i, u_i]$, target mass n_s, direction $d \in \{\text{lower}, \text{upper}\}$.
1: Set $w_i \leftarrow \ell_i$ and $R \leftarrow n_s - \sum_i \ell_i$.
2: If $R < 0$ or $R > \sum_i (u_i - \ell_i)$, return "incompatible envelope".
3: Sort indices by increasing y_i for lower, decreasing y_i for upper.
4: **for** indices i in sorted order **do**
5: $\quad \Delta_i \leftarrow \min\{R, u_i - \ell_i\}$; $w_i \leftarrow w_i + \Delta_i$; $R \leftarrow R - \Delta_i$.
6: $\quad$ Stop when $R = 0$.
7: **end for**
8: Return $\sum_i w_i y_i$.

Proposition J.1 (Correctness of Algorithm 3)**.** *When the stratum program is feasible, Algorithm 3 returns its exact minimum or maximum.*

Proof. We prove the minimum. Suppose a feasible solution has indices i, j with $y_i < y_j$, $w_i < u_i$, and $w_j > \ell_j$. Moving $\varepsilon = \min(u_i - w_i, w_j - \ell_j)$ units of mass from j to i preserves feasibility and decreases the objective by $\varepsilon(y_j - y_i) > 0$. Therefore an optimum cannot assign positive removable mass to a larger outcome while a smaller outcome has unused capacity. Repeated exchanges produce exactly the greedy allocation. The maximum follows by reversing the order. □

Some historical supplementary specifications permit a prespecified compatibility margin. The current semi-synthetic procedure keeps the frozen radius and basis and records a zero-dual fallback when training calibration is infeasible. The distinct historical routing rules are stated with their corresponding protocols below.

J.2 SEMI-SYNTHETIC MECHANISM-PROVENANCE PROTOCOL

Data, pairing, and the estimand. The fixed source is NVIDIA HelpSteer2 revision `990b2711a36180dd19d9c94b8627844866f8982a`, released under CC-BY-4.0. We group by exact prompt and require exactly two distinct responses, each matched by its original prompt/response key to the official train/validation records. Both arms must have at least two integer helpfulness and correctness ratings in $\{0, 1, 2, 3, 4\}$. Word count is the number of whitespace-separated tokens. Equal-length responses are excluded, and the shorter/longer responses become arms 0/1 together with their rating lists; no rating-based filtering is used. This gives $10{,}626$ pairs.

The prompt identifier is its UTF-8 SHA256 digest. Sorting by the digest of `DRIFT-real-v1-split|` concatenated with that identifier, breaking ties by identifier, fixes $2{,}125$ design, $2{,}125$ development, and $6{,}376$ confirmatory pairs. This experiment-specific HelpSteer2 split is defined for the present study. Design-only medians define four cells: prompt length at 39 words and response-length gap at 88 or 89.5 words within the low/high prompt groups, with equality assigned low. Confirmatory cell counts are $1609/1533/1539/1695$, so all population cells have positive mass.

Each historical trace samples a confirmatory pair uniformly with replacement. Let F_{ia} be the empirical distribution of arm a's ratings divided by four. The target averages $Y(1) - Y(0)$ over this finite corpus and fresh rating randomness. It compares complete responses, including their content differences, using existing responses and human ratings. The main outcome is helpfulness; correctness is a fixed secondary outcome.

Potential ratings and hidden selection. For each sampled pair draw $V \sim \mathrm{Unif}[0, 1)$ and set $Y(a) = F_{ia}^{-1}(V)$. In the associated design, $S = -1$ for $V < 0.5$ and $S = 1$ otherwise. The independent-state control draws a fair S independently of V. With cell probabilities $q = (0.30, 0.45, 0.55, 0.70)$ and signs $v = (-1, -1, 1, 1)$, the main historical selector is

$$p(i, S) = \mathrm{expit}\{\mathrm{logit}(q_j) + \eta(0.35v_j + 0.45S)\}.$$

Here η denotes experimental drift strength, distinct from the theoretical overlap bound. The exact maximal log-odds shift is 0.8η; thus the anchor $\eta = \Gamma = 0.6$ has actual maximal shift 0.48, within the analysis radius 0.6. The Γ sweep varies only the assumed radius while reusing historical/replay bundles. The η sweep changes the historical law at fixed $\Gamma = 1$.

The marginal-calibrated mechanism instead sets

$$p(i, S) = q_j + a_j S, \qquad a_j = 0.8 \min\{q_j - p_{Lj}^{(\eta)}, p_{Uj}^{(\eta)} - q_j\}, \quad p_{L/U,j}^{(\eta)} = \mathrm{expit}\{\mathrm{logit}(q_j) \mp \eta\}.$$

The fair state gives $\mathbb{E}[p \mid Z] = q$ for both associated and independent-state controls, while association changes its coupling with the ratings. This equality does not generally hold for the independent-state logit-shift control. Reverse-rank coupling uses $Y(0) = F_{i0}^{-1}(1-V)$; the inverse empirical CDF clips its endpoint index to the last rating. A separate uniform draw generates $D \sim \mathrm{Bernoulli}(p(i, S))$ and reveals only $Y = DY(1) + (1 - D)Y(0)$.

Replay and calibration. Total replay budget B is allocated equally, $B_j = B/4$, across all four population cells independently of the complete historical sample. Cellwise Clopper–Pearson intervals use failure budget $\alpha_{\mathrm{rep}}/4$, with $\alpha_{\mathrm{rep}} = 0.01$. Each interval endpoint is projected into the prespecified known selector range $[0.30, 0.70]$. Endpointwise projection preserves simultaneous coverage because every true q_j belongs to that range. The log-odds bridge then defines the weight envelopes for every cell, including cells absent from a particular history. Known-q conditions supply the exact q_j to the DRIFT envelope.

Calibration uses nested groups: $K = 1$ is an intercept; $K = 2$ is an intercept plus the high-prompt indicator; $K = 4$ is an intercept plus three cell indicators. They are equivalently optimized using one threshold per group. The default threshold grid is $\{0, 0.25, 0.5, 0.75, 1\}$, the full rating support; ties select the smallest threshold. The approximate-training variant fixes $\{0, 0.5, 1\}$.

Held-out scores and finite-sample correction. Two folds are fixed by record index modulo two. Each fold's four dual objectives use only its complementary historical fold, with the independent replay envelope held fixed. Empirical box/moment incompatibility or an exception during threshold

fitting is recorded and invokes the zero-dual fallback. Unhandled outer-run exceptions are retained as failed records. No historical-data-dependent radius enlargement or moment relaxation is allowed.

For every returned threshold and replay realization,

$$\bar{w} = 1 + \tfrac{7}{3}e^{\Gamma}, \qquad g_a^{\pm} \in [1-\bar{w}, \bar{w}], \qquad |g_a^{\pm} - 0.5| \le B_* := \bar{w} - 0.5.$$

The implementation applies (70) to these centered scores with $\delta = (0.05 - 0.01)/(4 \times 2) = 0.005$, retaining both variance and range terms. Centering changes neither variance nor the deviation of the mean from its expectation, so the stored means of g are corrected directly. Scores are checked against these deterministic bounds. Aggregation weights equal evaluation-fold size divided by n.

Twenty positive-radius conditions implement the finite-sample construction under its noncollapsed boxes. Two known-q, $\Gamma = 0$ conditions are separate compatible boundary checks: the boxes collapse, strict interiority is unavailable, and exact LCE identification plus bounded held-out concentration supplies the direct justification. Each repetition uses four expected zero-dual fold/arm fallbacks there; positive certificate width reflects historical sampling rather than positive-radius ambiguity.

Comparators and invalid ablations. Replay-IPW substitutes the cell replay mean clipped to $[0.15, 0.85]$ into the local IPW score. Cell-AIPW uses two-fold complementary-history estimates: the smoothed cell propensity is $(S_j + 0.5)/(N_j + 1)$, where N_j is the complementary-training-cell count and S_j its arm-1 count, distinct from the hidden state S. This estimate is projected between the global bridge-implied historical probability limits, and each arm mean is the selected training-cell mean. An empty arm/cell falls back to the training-fold arm mean, then 0.5 if that arm is absent entirely. These nuisance functions enter the usual held-out augmented IPW score. Both methods use ordinary inputs only; historical-p IPW is explicitly an oracle reference. Known-q DRIFT does not change replay-IPW's use of $\widehat{q}$.

The three ablations respectively replace the replay band by a projected $\widehat{q}$ point, remove both Bernstein terms, or fit on the full history while evaluating on its two folds. Each removes a condition used for finite-sample coverage. The primary empirical event is the literal complete chain; scalar τ coverage and each set-containment link are also recorded. Every frozen row has $\tau \in \mathcal{I}$, so the stored conjunction of the two set links equals the reported literal chain. Empty intervals fail the chain check.

Freeze, oracle boundary, and numerical checks. The formal matrix has 15 main, three outcome/coupling, and seven method-variant conditions, with $1{,}000$ repetitions each under root seed `20260915`. Separate PCG64 streams generated by SeedSequence govern pair sampling, rating rank, independent-state control, historical selection, and replay. Per-condition seeds and bundle identifiers are saved. Whole-data structural and rating-list checks preceded the freeze; the formal effects, coverage, and widths did not select the frozen conditions.

Exact finite (j, D, Y) population laws and endpoints are precomputed for evaluation. Ordinary estimator functions neither accept those references nor inspect hidden mechanisms: ordinary outputs are sealed before evaluator-only calculations are used in each repetition. HiGHS primal programs and threshold-dual solutions independently agree to maximum 9.99×10^{-15}; the population historical-IPW identity agrees to 8.26×10^{-16}. The width decomposition agrees to 5.56×10^{-16}. In the two marginal-calibrated $K = 4$ conditions, a range-matched quantile-balancing comparison (Dorn & Guo, 2023) restricts context to the same four-cell partition, where $e(j) = q_j$, and uses odds factor $\Lambda = \exp(\Gamma)$. Its population endpoints agree to 3.33×10^{-16}. This numerical check establishes endpoint agreement under the shared four-cell context model. Full-pair conditioning and continuous-outcome bootstrap inference require separate analyses.

The three diagnostic tables above are deterministic post-hoc summaries of the frozen records, specified after the confirmatory run. For an inner interval $A = [L_A, U_A]$ and intended outer interval $B = [L_B, U_B]$, containment uses $L_B \le L_A + 10^{-10}$ and $U_A \le U_B + 10^{-10}$, and the reported violation is $\max\{L_B - L_A, U_A - U_B, 0\}$. The stored `full_chain` field joins the two set-containment links; because all $25{,}000$ records also satisfy $\tau \in \mathcal{I}$, it is numerically identical here to the literal chain $\tau \in \mathcal{I} \subseteq \widetilde{\mathcal{I}} \subseteq \mathcal{C}$. Single-condition MCSEs use the $1{,}000$ repetitions. Comparisons among the anchor, Γ/K variants, known-q, coarse-grid, and ablation rows align their shared `sample_bundle_id`; conditions with different n or B use distinct bundles. The $25{,}000$ rows contain $13{,}000$ distinct bundles and are never treated as independent repetitions of one pooled design.

Table 7: All 25 frozen conditions, each with 1,000 completed repetitions and no failed or pending runs. n and B are in thousands; widths are normalized rating-difference units. Chain is the complete-set event frequency; scalar $\tau \in \mathcal{C}$ is 1.000 for every row, including invalid ablations. H/C denote helpfulness/correctness. The two zero-drift rows are compatible boundary checks; the three invalid rows carry no theorem-backed coverage promise.

Condition	η	Γ	K	$n/10^3$	$B/10^3$	Chain	$\lvert\mathcal{I}\rvert$	$\lvert\mathcal{C}\rvert$
Logit shift	0.6	0.6	4	2	1	1.000	0.267	1.183
Logit shift	0.6	0.6	4	2	10	1.000	0.267	1.110
Logit shift	0.6	0.6	4	10	1	1.000	0.267	0.590
Logit shift (anchor)	0.6	0.6	4	10	10	1.000	0.267	0.522
Logit shift	0.6	0.6	4	50	1	1.000	0.267	0.441
Logit shift	0.6	0.6	4	50	10	1.000	0.267	0.374
Logit shift	0.6	0.8	4	10	10	1.000	0.363	0.635
Logit shift	0.6	1.0	4	10	10	1.000	0.454	0.744
Logit shift	0.0	1.0	4	10	10	1.000	0.472	0.762
Logit shift	0.3	1.0	4	10	10	1.000	0.467	0.758
Logit shift	1.0	1.0	4	10	10	1.000	0.426	0.722
Independent state (logit)	0.6	0.6	4	10	10	1.000	0.269	0.525
Marginal, associated	0.6	0.6	4	10	10	1.000	0.294	0.540
Marginal, independent	0.6	0.6	4	10	10	1.000	0.299	0.546
Zero drift, known q (H)	0.0	0.0	4	10	10	1.000	0.000	0.242
Reverse-rank coupling	0.6	0.6	4	10	10	1.000	0.261	0.515
Correctness	0.6	0.6	4	10	10	1.000	0.258	0.513
Zero drift, known q (C)	0.0	0.0	4	10	10	1.000	0.000	0.246
$K = 1$	0.6	0.6	1	10	10	1.000	0.299	0.550
$K = 2$	0.6	0.6	2	10	10	1.000	0.268	0.522
Known q	0.6	0.6	4	10	10	1.000	0.267	0.486
Coarse threshold grid	0.6	0.6	4	10	10	1.000	0.267	0.539
Plug-in $\widehat{q}$ (invalid)	0.6	0.6	4	10	10	0.146	0.267	0.482
No Bernstein (invalid)	0.6	0.6	4	10	10	0.259	0.267	0.305
Overlapping sample (invalid)	0.6	0.6	4	10	10	1.000	0.267	0.517

Table 8: Link-specific diagnostics for the anchor and invalid ablations. Each entry is successes/1,000, followed by its pointwise Wilson 95% Monte Carlo interval. These are descriptive checks of the frozen records; invalid ablations carry no coverage promise.

Method	Literal chain	$\mathcal{I} \subseteq \widetilde{\mathcal{I}}$	$\widetilde{\mathcal{I}} \subseteq \mathcal{C}$	$\mathcal{I} \subseteq \mathcal{C}$	$\tau \in \mathcal{C}$
Full	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]
Point $\widehat{q}$	146/1000 [0.125,0.169]	146/1000 [0.125,0.169]	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]
No Bernstein	259/1000 [0.233,0.287]	1000/1000 [0.996,1.000]	259/1000 [0.233,0.287]	955/1000 [0.940,0.966]	1000/1000 [0.996,1.000]
Same sample	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]	1000/1000 [0.996,1.000]

Table 9: Post-hoc magnitude and precision diagnostics from the frozen records. Violation summaries condition on violations above 10^{-10} while retaining their counts. Interval summaries use all 1,000 anchor repetitions. MC intervals quantify simulation precision.

Containment link	Positive/1,000	Positive median	Positive 95% quantile	Maximum
Point $\widehat{q}$: $\mathcal{I} \to \widetilde{\mathcal{I}}$	854/1000	0.00344	0.00916	0.01378
No Bernstein: $\widetilde{\mathcal{I}} \to \mathcal{C}$	741/1000	0.00604	0.01831	0.02954
No Bernstein: $\mathcal{I} \to \mathcal{C}$	45/1000	0.00233	0.00942	0.01610

Anchor quantity	Mean	MCSE	5%	Median	95%
$\widetilde{\mathcal{I}}$ lower	-0.08817	0.00007	-0.09153	-0.08836	-0.08401
$\widetilde{\mathcal{I}}$ upper	0.21432	0.00007	0.21076	0.21435	0.21797
$\widetilde{\mathcal{I}}$ width	0.30249	0.00011	0.29659	0.30262	0.30782
$\mathcal{C}$ lower	-0.19975	0.00030	-0.21470	-0.19974	-0.18414
$\mathcal{C}$ upper	0.32208	0.00027	0.30797	0.32215	0.33593
$\mathcal{C}$ width	0.52183	0.00030	0.50614	0.52179	0.53791

Table 10: Signed Monte Carlo mean errors in the marginal-calibrated control. Pointwise MC intervals quantify simulation precision across 1,000 frozen repetitions.

Hidden association	Method	Mean error	MCSE	Pointwise MC 95% interval
Independent	Replay-IPW	-0.000508	0.000693	[-0.001868,0.000852]
Independent	Cell AIPW	0.000295	0.000213	[-0.000122,0.000712]
Associated	Replay-IPW	0.047719	0.000679	[0.046387,0.049051]
Associated	Cell AIPW	0.047414	0.000212	[0.046998,0.047829]

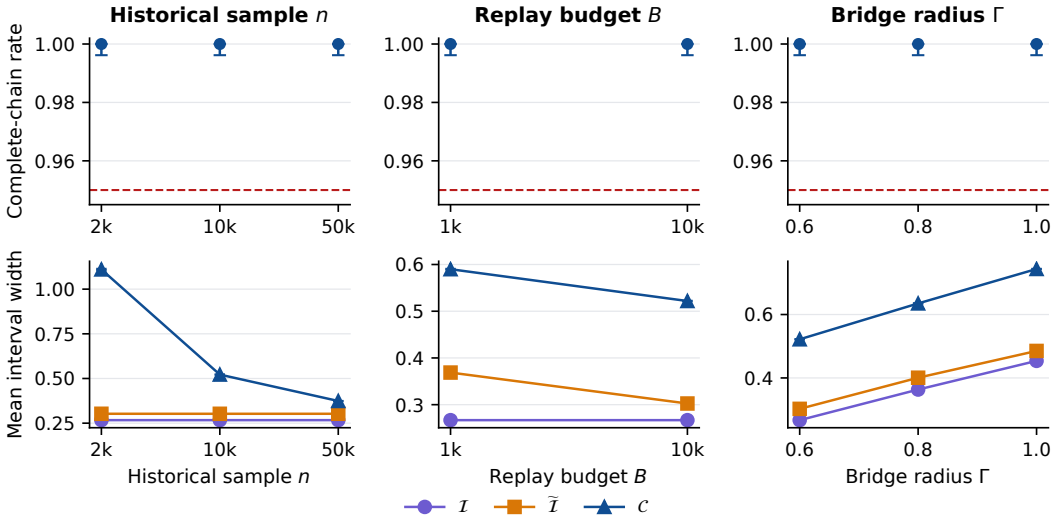

Figure 12: **Complete-chain frequency and width across the frozen parameter grid.** This post-hoc display summarizes the frozen records. All panels use the common helpfulness, logit-shift, hidden-association, shared-rank, `drift_full` specification satisfying the coverage assumptions. The top row reports the literal event $\tau \in \mathcal{I} \subseteq \widetilde{\mathcal{I}} \subseteq \mathcal{C}$: every plotted condition has $1{,}000/1{,}000$ successes, with pointwise Wilson 95% Monte Carlo interval $[0.9962, 1]$ and the nominal 0.95 reference. The intervals quantify pointwise simulation precision. An observed rate of 1.000 corresponds to 100% success in the simulated repetitions; population coverage follows from the theorem conditions. The bottom row reports mean widths of $\mathcal{I}$, $\widetilde{\mathcal{I}}$, and the outer certificate $\mathcal{C}$ in normalized rating-difference units, with pointwise 95% Monte Carlo intervals. The n panel fixes $(B, \Gamma, \eta, K) = (10{,}000, 0.6, 0.6, 4)$; the B panel fixes $(n, \Gamma, \eta, K) = (10{,}000, 0.6, 0.6, 4)$; and the Γ panel fixes $(n, B, \eta, K) = (10{,}000, 10{,}000, 0.6, 4)$. The Γ levels share bundles within replicate, whereas the n and B comparisons use distinct bundles and are not treated as paired; no conditions are pooled. The separate nominal 0.99 replay-band population event is not plotted.

J.3 Supplementary Experiment 4: Prespecified drift-robustness protocol

The final six-stratum design uses $n = 600$, stratified replay budget $B = 3000$, 400 repetitions at each $\Gamma \in \{0, 0.15, 0.30, 0.45, 0.60\}$, and root seed 20260911. The following definitions reproduce the finite generated-pair implementation. For zero-indexed strata $j = 0, \ldots, 5$, let $d_j = (-1)^j$, $z_j = 2j/5 - 1$, and $\mathrm{expit}(t) = (1 + e^{-t})^{-1}$. The current and historical display probabilities are

$$q_j = 0.24 + 0.52j/5, \qquad \gamma_j = \Gamma\{0.65 + 0.35(j+1)/6\}, \qquad p_j = \mathrm{expit}\{\mathrm{logit}(q_j) + d_j\gamma_j\}.$$

Stratum masses are $(0.24, 0.20, 0.18, 0.15, 0.13, 0.10)$. Independently of the stratum, $X_1 \sim N(0,1)$ and $X_2 \sim \mathrm{Unif}(-1, 1)$; conditional on them, $U \sim \mathrm{Bernoulli}\{\mathrm{expit}(0.4X_1 - 0.25X_2)\}$. Write $b_j = \mathrm{expit}(-1.15 + 0.38z_j)$, $t_j = 0.09 + 0.08j/5 + 0.035d_j$, and $\mathrm{clip}_{[a,b]}(v) = \min\{b, \max(a, v)\}$. The conditional outcome means are

$$m_0 = \mathrm{clip}_{[0.03,0.86]}\{b_j + 0.035\tanh(X_1) + 0.025U\},$$
$$m_1 = \mathrm{clip}_{[0.04,0.96]}\{m_0 + t_j + 0.025\tanh(X_2) - 0.02U\}.$$

Potential outcomes are Bernoulli draws with these means, and $D \sim \mathrm{Bernoulli}(p_j)$ is generated by a separate stream. The observed outcome is $Y = DY(1) + (1-D)Y(0)$. The repetition-specific causal truth is the mean of $m_1 - m_0$; these means, U, and true propensities remain evaluation-only inputs.

The simultaneous exact-binomial replay band has error allocation 0.02 and overlap floor 0.05. The base certified radius is $\Gamma + 0.08$. The smallest additional nonnegative inflation δ_{feas} making the empirical stratum–arm normalizations feasible is determined using only (D, j, q_L, q_U), with bisection tolerance 10^{-10}, numerical pad 10^{-9}, and maximum additional radius 3. Outcomes, causal truth, and width are excluded from calibration. The final radius is $\Gamma + 0.08 + \delta_{\mathrm{feas}}$. Historical oracle IPW remains an oracle benchmark; current replay IPW, replay AIPW, box-only sensitivity, and calibrated DRIFT use their specified observed inputs. Truth containment, identified width, finite-sample outer width, and Hausdorff endpoint error are reported separately.

Across the increasing drift grid, mean δ_{feas} is 0.078073, 0.038036, 0.028945, 0.028782, and 0.029848; the corresponding fractions with positive inflation are 0.5550, 0.2875, 0.2475, 0.2150, and 0.2050. These values quantify the outcome-blind adjustment required for empirical compatibility.

J.4 SUPPLEMENTARY EXPERIMENT 5: PRESPECIFIED MOMENT-RESOLUTION PROTOCOL

The final run uses the finite-stratum law above at $\Gamma = 0.6$, $n = 600$, $B = 3000$, and 400 repetitions with root seed 20260913 and namespace $(5, 3, r)$. Every repetition reuses one observed bundle across $K \in \{1, 3, 5, 6\}$. With $z_1 = X_2$ and $z_2 = 2j/5 - 1$, the nested bases are (1), $(1, z_1, z_2)$, $(1, z_1, z_2, z_1^2, z_2^2)$, and $(1, z_1, z_2, z_1^2, z_2^2, z_1 z_2)$. Two folds are assigned by row index modulo two before outcome access.

The certificate fixes $\alpha = 0.10$, replay allocation $\alpha_{\mathrm{rep}} = 0.02$, and held-out component allocation $(\alpha - \alpha_{\mathrm{rep}})/(4\times 2) = 0.01$. Exact-binomial simultaneous replay bands are inflated by the prespecified drift radius. The causal endpoints combine the arm endpoints as $[L_1 - U_0, U_1 - L_0]$. HiGHS solves each empirical program; primal residual, endpoint-ordering, and primal–dual agreement tolerances are 10^{-8}. Solver-certified infeasibility invokes the prespecified zero-dual certificate on that fold–arm component, while unknown, unbounded, numerical-error, exception, or inconsistent states block execution. There is no outcome-driven moment slack or radius enlargement. All $1{,}600$ repetition–K rows, including conservative fallback components, remain in the analysis. Population endpoints are opened only after ordinary estimates and routing decisions have been fixed.

J.5 EXP6: JOINT HISTORICAL-REPLAY RATE PROTOCOL

The final joint-rate design fixes $\Gamma = 0.6$ and $K = 3$. Historical sample sizes are $n \in \{1200, 2400, 4800\}$; replay budgets are $B \in \{240, 480, 960, 1920\}$. Each cell contains 400 repetitions. Root seed 20260920 uses namespace $(6, 4, r)$; cells within a repetition share the parent namespace, with independent historical and replay child streams. The bounded continuous maximum-entropy outcome generator preserves the finite-stratum conditional means and removes outcome atoms at dual kinks. Two row-index folds and the basis $(1, X_2, 2j/5 - 1)$ are fixed before outcomes.

The ordinary replay estimate is globally fixed to $\widetilde{q}_j = (S_j + 1/2)/(B_j + 1)$, without adaptive clipping. The primary empirical dual is solved with CLARABEL at radius $R_K = 10$ and checked independently using SLSQP with objective tolerance 10^{-6}. Population integration uses production order 24 and audit order 16. Population endpoints are evaluation-only. Before forming effect endpoints as lower-arm-1 minus upper-arm-0 and upper-arm-1 minus lower-arm-0, the saved estimator replaces any crossed lower/upper estimates within an arm by their common midpoint. This arm-order projection occurs in $132/4{,}800$ repetitions; the reported table and heatmaps use the projected estimates. The unprojected comparison and maximum endpoint change are reported in Section F.3.6. The no-intercept nonnegative least-squares fit of projected-endpoint MSE to $C_1/n + C_2/B$ is supplemented by an unconstrained fit and a repetition-grouped bootstrap with $10{,}000$ resamples, seed 2026092006, and 95% intervals. Coefficient signs, fit quality, and monotonicity are diagnostics rather than execution gates; the full $4{,}800$ repetitions are retained.

J.6 Supplementary Experiment 7: prespecified optimized-endpoint inference protocol

The final design fixes $\Gamma = 0.35$, $K = 3$, two balanced folds, dual radius 10, $n = B = 21{,}120$, and $1{,}000$ repetitions under root seed 2026090510 and namespace $(7, 4, r)$. It uses the bounded continuous maximum-entropy outcome law and the protected optimized-endpoint estimator. Independent finite-stratum replay uncertainty is added to the joint historical-score covariance for studentization, as in (83); population endpoint truth enters bias and coverage evaluation only.

All saved repetitions are checked for two-sample covariance reconstruction and historical/replay seed separation. Primary dual fits receive the independent SLSQP check throughout. Exact primal–dual LP comparisons additionally cover prespecified repetitions 0 and 999, or 16 fits. The repetition count was fixed prospectively, and all completed outcomes are retained. Active-branch disagreement, centering, scale, and tail shape are reported as diagnostics.

J.7 Supplementary Experiment 8: prespecified finite-pilot audit protocol and complete results

Archive scope. The following protocol, tables, and figures reproduce the historical EXP8 report. The release retains archived source code and 282 executed-prefix files from 400 reported repetitions, but lacks the complete row-level results and pilot records required to independently recompute all summaries. These are restored archived summaries with partial raw evidence, not a fresh execution or full numerical revalidation.

The final binary-outcome design uses the finite-stratum law of Section J.3, with $\Gamma = 0.6$, $n = 12{,}000$, $J = K = 6$, and 400 repetitions under root seed 2026090582. Every cell receives 4000 pilot historical-mechanism draws. Replay-to-audit count ratios are $\beta = (1, 2, 4, 1, 2, 4)$ and historical-audit unit costs are $\kappa = (0.7, 1, 1.4, 1.9, 2.5, 3.2)$. Thus the pilot costs 42,800. Replay and historical-trace acquisition have zero incremental cost in this experiment; replay draws are nevertheless counted. Main budgets are 12,000, 24,000, 48,000, and 96,000.

For m_j historical and $b_j = \beta_j m_j$ replay draws, the audit certificate uses $\alpha_{\mathrm{aud}} = 0.05$ and Hoeffding half-widths

$$\left\{\frac{\log(4J/\alpha_{\mathrm{aud}})}{2m_j}\right\}^{1/2} \quad \text{and} \quad \left\{\frac{\log(4J/\alpha_{\mathrm{aud}})}{2b_j}\right\}^{1/2}.$$

Probability-band endpoints are clipped at $\eta_0 = 0.025$ and $1 - \eta_0$. The cell drift bound is the maximum absolute logit difference across opposite historical/replay band endpoints. This bound and the replay band define the empirical propensity envelope. Each arm endpoint is computed by the stratified linear program and independently checked by continuous knapsack, with primal–dual agreement, equality residual, and box-violation tolerance 10^{-8}. These empirical intervals do not include the held-out outer-confidence-set correction.

Pilot sensitivities are one-sided width differences with step 0.001 at the pilot upper-radius reference, holding the pilot replay band fixed. Let $s_j^+ = \max(\widehat{s}_j, 0)$ and

$$c_j = \frac{2}{\eta_0(1-\eta_0)}\sqrt{\frac{\log(4J/\alpha_{\mathrm{aud}})}{2}}(1 + \beta_j^{-1/2}).$$

Continuous main allocations have relative weights κ_j^{-1} (Uniform), $(c_j/\kappa_j)^{2/3}$ (Uncertainty-only), $(s_j^+/\kappa_j)^{2/3}$ (Leverage-only), and $(s_j^+ c_j/\kappa_j)^{2/3}$ (Full plug-in), under the same total-cost constraint and a minimum of 200 historical draws per cell. All-zero allocation weights invoke equal-cost allocation. Integer counts are obtained by flooring and a single pass in descending fractional-remainder order, breaking ties by cell index and adding a draw only if affordable. History and pilot are shared across policies; independent main streams supply common prefixes. Plans are saved before those main draws are revealed.

All planned outcomes are retained. At the largest budget, widths are summarized only over feasible programs; an unconditional mean is unavailable whenever any outcome is undefined. Paired effects use the jointly feasible subset and pointwise unadjusted t Monte Carlo intervals. Neither undefined widths nor missing pairs are zero-filled.

Table 11: Complete Experiment 8 width and cost summary. Every row retains all 400 planned outcomes; n_f is the feasible denominator. Width and MCSE are conditional on feasibility. C_{all} and C_f are actual total costs averaged over all 400 outcomes and over the same feasible subset as width, respectively, including pilot cost. At the first three budgets conditional and unconditional widths coincide; at $96{,}000$ unconditional widths are undefined. Costs are shown in thousands.

Main budget	Method	$n_f/400$	Mean width	MCSE	$C_{\text{all}}/10^3$	$C_f/10^3$
12,000	Uniform	400/400	0.438179	0.000893	54.799800	54.799800
12,000	Uncertainty-only	400/400	0.439482	0.000928	54.800000	54.800000
12,000	Leverage-only	400/400	0.443913	0.000945	54.799584	54.799584
12,000	Full plug-in	400/400	0.441890	0.000917	54.799551	54.799551
24,000	Uniform	400/400	0.339074	0.000812	66.799800	66.799800
24,000	Uncertainty-only	400/400	0.340280	0.000798	66.799100	66.799100
24,000	Leverage-only	400/400	0.341502	0.000825	66.799571	66.799571
24,000	Full plug-in	400/400	0.340002	0.000823	66.799568	66.799568
48,000	Uniform	400/400	0.259012	0.000880	90.799900	90.799900
48,000	Uncertainty-only	400/400	0.260338	0.000873	90.799000	90.799000
48,000	Leverage-only	400/400	0.259933	0.000876	90.799577	90.799577
48,000	Full plug-in	400/400	0.258895	0.000882	90.799565	90.799565
96,000	Uniform	399/400	0.190837	0.001012	138.799700	138.799700
96,000	Uncertainty-only	398/400	0.192072	0.001000	138.799300	138.799300
96,000	Leverage-only	396/400	0.193047	0.000983	138.799538	138.799534
96,000	Full plug-in	397/400	0.192257	0.000980	138.799525	138.799524

Table 12: All Experiment 8 Full-minus-baseline paired width contrasts. n_p counts jointly feasible repetitions out of 400. Negative effects favor Full plug-in. Pointwise, unadjusted 95% t intervals quantify Monte Carlo uncertainty in the paired width contrasts.

Main budget	Baseline	$n_p/400$	Difference	MCSE	95% MC interval
12,000	Uniform	400/400	0.003712	0.000336	$[0.003051, 0.004372]$
12,000	Uncertainty-only	400/400	0.002408	0.000355	$[0.001709, 0.003107]$
12,000	Leverage-only	400/400	-0.002023	0.000189	$[-0.002394, -0.001651]$
24,000	Uniform	400/400	0.000928	0.000276	$[0.000386, 0.001471]$
24,000	Uncertainty-only	400/400	-0.000278	0.000238	$[-0.000746, 0.000189]$
24,000	Leverage-only	400/400	-0.001500	0.000145	$[-0.001785, -0.001215]$
48,000	Uniform	400/400	-0.000117	0.000213	$[-0.000535, 0.000302]$
48,000	Uncertainty-only	400/400	-0.001442	0.000205	$[-0.001845, -0.001039]$
48,000	Leverage-only	400/400	-0.001037	0.000120	$[-0.001273, -0.000802]$
96,000	Uniform	397/400	0.001263	0.000184	$[0.000902, 0.001624]$
96,000	Uncertainty-only	397/400	0.000174	0.000171	$[-0.000162, 0.000509]$
96,000	Leverage-only	396/400	-0.000722	0.000107	$[-0.000932, -0.000511]$

The deterministic population comparison appears once, as panels C–D of Figure 11. The first panel compares exact population width $W(\Gamma_\infty + c/\sqrt{m})$ with the first-order objective at the population first-order allocation; the second compares analytic sensitivities with one-sided finite differences at the true radius using step 10^{-5}. Large auxiliary budget parameters index this deterministic comparison; actual audit expenditures and simulation conditions are reported separately. The approximately 0.335% gain over equal cost is the reduction in the first-order theoretical objective. Finite-budget remainders can separate the first-order objective from exact width. Consistency of the practical finite-pilot sensitivities at Γ_∞ remains an open question.

J.8 Open Bandit Dataset protocol

We use the 10,000-row public quick-start files for the "all" campaign under the Bernoulli Thompson-sampling and uniform random policies. Each record contains the displayed item, click reward, position, logged behavior propensity, hashed user features, and user–item affinity vector. The companion item table supplies one numerical and three categorical item attributes. The target policy is uniform over the 80-item action set, so $\pi(a \mid z) = 1/80$.

This finite-action stress test evaluates replay-envelope coverage, overlap, and contraction under audited historical probabilities using real logged contexts and rewards. The eligible-pair target τ_h requires a fixed binary candidate pair before display, which is unavailable in this dataset.

The historical finite-action propensity in the Thompson log is treated as hidden from every nonoracle method. Equation (113) generates the current replay score from the real historical score, item attribute, time, and position. This is a controlled stress test: the rewards, contexts, displayed actions, and historical propensities are real; only the version drift is injected. The randomized-policy click mean is an on-policy benchmark rather than the exact finite-sample value of the Thompson-log importance-weighted sample, so discrepancies remain even for oracle IPS when effective sample size is low.

For the auxiliary finite-action calculation, the nominal target importance weight is $\pi(A_i \mid Z_i)/q(A_i \mid Z_i)$, and (130) supplies the multiplicative box. The described implementation enforces exact weight normalization within each observed context stratum. If the box is incompatible with that normalization, it checks log-radius inflations from 0 to 3.0 in increments of 0.025 and reports failure if none is feasible. It optimizes both endpoints on the same sample and expands each endpoint by 1.96 times the empirical standard deviation of its optimized weighted rewards divided by $\sqrt{n}$. These sampling expansions give diagnostic intervals. The certificate in Theorem D.3 additionally requires an independent population-valid replay band. Bridge audits reveal historical propensities for budgeted rows. For a stratum with m_j audited rows, the implemented radius is the smaller of the global radius and the empirical 98th percentile of the absolute audited log-ratios plus $0.10 + 1.5/\sqrt{m_j}$; a stratum with no audited row uses the global radius. A population-uniform bridge would require additional justification beyond these empirical quantiles and sensitivity margins.

Table 13: Descriptive quantities for the referenced 10,000-row quick-start files.

Quantity	Thompson log	Random log
Interactions	10,000	10,000
Distinct items	80	80
Observed click rate	0.0042	0.0038
Mean logged propensity	0.1089	0.0125
Minimum logged propensity	4.5×10^{-5}	0.0125
Maximum logged propensity	0.9542	0.0125
Uniform-policy oracle IPS from Thompson log	0.00236	N/A
Uniform-policy oracle effective sample size	340.4	10,000

J.9 Additional numerical summaries and artifact availability

The companion experiment-source bundle contains the HelpSteer2 study in `HelpSteer2/`. It supplies `formal_protocol.yaml`, the historical pre-run manifest under `provenance/`, a text-free `data/prepared/pairs_runtime.parquet`, reference ratings, the $25{,}000$-row `results/formal/raw_results.parquet`, `summary.csv`, population diagnostics, an artifact-level validator, and figure scripts. The three upstream JSONL files and the derived prompt/response columns are not redistributed. Their fixed public revision, expected hashes, and deterministic fetch/preparation scripts are provided for optional reconstruction. The complete formal matrix is listed in Table 7. Publication figures read only saved result summaries.

The package also supplies `simulation_exp01_exp07/data/`, `results/`, `figures/`, and `paper_illustrations/` as retained EXP1–7 evidence. The historical runner framework is not distributed, so these diagnostics are not end-to-end reproducible from the release. Public Open Bandit input CSVs are also omitted; the package README identifies the retained Open-Bandit-specific outputs.

J.10 Reproducibility commands and boundaries

From `experiment_source/`, `python verify_release.py` checks the release manifest, path safety, anonymization, privacy-minimized runtime corpus, and retained $25{,}000$-row primary result table. From `HelpSteer2/`, `python verify_formal_results.py` performs artifact-level structural and numerical checks, and `python -m unittest discover -s tests -v` checks the implementation components. `python make_publication_figures.py` regenerates publication assets from saved summaries. Optional corpus reconstruction uses `python fetch_helpsteer2.py` followed by `python prepare_data.py`. `EVIDENCE_MAP.md`

records which claims rely on reproducible HelpSteer2 artifacts and which rely only on retained outputs.

The manuscript compiles from the submission-source directory using `latexmk -pdf -interaction=nonstopmode -halt-on-error main.tex`. The EXP1–7 historical runner framework is outside the anonymous release; their reported results refer to retained outputs. The separate `EXP8/` directory retains the reported protocol and tables, original figures, historical source, and partial executed-prefix data; its README lists missing full-run artifacts. Artifact integrity and execution reproducibility are distinct checks.

K Evidence Map and Four-Question Results

Table 14: Related models and their connections to the historical/current observation scheme and mechanism-evidence objective.

Setting	Uncertainty and evidence	Target and connection
LCE (Athey et al., 2026)	Relative historical selection; reproducible resampling when needed	Characterizing contrast inherited; unavailable historical selector motivates a bridge
Sharp sensitivity (Dorn & Guo, 2023; Ishikawa & He, 2023; Zhang & Zhao, 2026)	Confounding restrictions and observational outcomes	Established sharp-bound/moment tools; our anchor is current replay
Robust OPE (Bennett et al., 2024; Hess et al., 2026)	Transition perturbations or hidden confounding	Efficient sharp policy-value bounds in these models; DRIFT efficiency remains uncharacterized
Structured OPE (Khan et al., 2024)	Outcome smoothness supplies information beyond overlap	Structured scalar-target inference uses information beyond the unstructured surface model
Bound-directed design (Ailer et al., 2024; Maringgele & Etesami, 2026)	Indirect experiments or costly experiment subsets narrow causal bounds	Width-directed acquisition; here, evidence consists of historical-selector audits
This paper	Historical outcomes, separate current replay, and historical mechanism audits	Marginal eligible-pair width and its audit derivative; latent effects receive outer validity

The eight synthetic diagnostics and the frozen HelpSteer2 study address the four questions in Section 7 under distinct designs. EXP1–8 retain their original controlled laws and saved outputs, but their historical runner is not distributed. HelpSteer2 uses one fixed corpus with constructed selectors and potential-rating coupling shared across its finite-sample, marginal-association, calibration, and budget comparisons. The package-level `EVIDENCE_MAP.md` links each claim to its retained artifact and reproducibility status.

Estimator inputs define the ordinary/evaluator boundary. Ordinary functions receive selected ratings, logged identifiers/cells, and independent replay data. Exact population quantities, hidden state, potential ratings, and historical propensities remain evaluator-only. The companion supplies a privacy-minimized runtime corpus, reference ratings, the $25{,}000$-row primary result table, hashes, summaries, source code, and numerical validators. Text-bearing upstream files are reconstructed from the pinned public revision and are not redistributed.

K.1 HelpSteer2 validation across the four questions

1. **Missing historical coupling.** The marginal-calibrated controls hold $\mathbb{E}[p \mid Z] = q$ exactly while switching the hidden state–rating association. Point bias changes under marginal selector agreement, illustrating the mechanism in Theorem B.2. The theorem supplies the separate observational-equivalence result.
2. **Marginal sets and compatible zero drift.** The K variants and paired Γ sweep compare admissible weight restrictions at a fixed corpus. Known-q, zero-drift rows recover singleton

population sets while retaining nonzero finite-sample widths. Their collapsed boxes are treated as boundary checks using LCE and concentration rather than strict-interiority instances.

3. **Finite-sample certification.** The 20 positive-radius conditions combine all-population-cell replay bands, fixed folds, bounded held-out scores, and the full empirical-Bernstein correction. All $1{,}000$ literal chains $\tau \in \mathcal{I} \subseteq \widetilde{\mathcal{I}} \subseteq \mathcal{C}$ hold in each condition. The two zero-drift checks also have $1{,}000/1{,}000$ chains. Individual Wilson intervals quantify pointwise simulation precision.
4. **Implementation conditions.** Post-hoc diagnostics of the frozen records identify the failed inclusion links. The plug-in-$\widehat{q}$ variant has $\mathcal{I} \subseteq \widetilde{\mathcal{I}}$ in 0.146 of repetitions but directly contains $\mathcal{I}$ in every final $\mathcal{C}$. Removing Bernstein corrections has $\widetilde{\mathcal{I}} \subseteq \mathcal{C}$ in 0.259 and direct $\mathcal{I} \subseteq \mathcal{C}$ in 0.955; all scalar truths nevertheless remain covered. Positive-violation magnitudes and their full denominators are reported separately. Overlapping training/evaluation has favorable empirical events without the theorem's independence guarantee. Coarse-grid training retains valid score construction with a wider certificate.
5. **Sampling versus identification.** The $n \times B$ comparisons, endpoint distributions, and width decomposition distinguish historical concentration, replay enlargement, and fixed-radius population ambiguity. The anchor population set and all final certificates include zero. Pointwise MC intervals use the $1{,}000$ condition repetitions; shared-bundle method/radius comparisons are paired, while distinct n/B bundles are not. The regular two-sample expansion and minimax rate are assessed separately.
6. **Historical-evidence boundary.** The Γ sweep measures the width implications of different admissible bridges. It does not evaluate the first-order audit-allocation theorem. EXP8 separately reports finite-pilot audit rankings.

K.2 Synthetic evidence across the four questions

EXP1–3 address where information lives: expanding exact-treatment support versus a scalar local contrast, algebraic zero-drift recovery, and the strict two-world construction. EXP4–5 distinguish population calibration from finite-sample certificate complexity. EXP6–7 address historical/replay uncertainty: the descriptive EXP6 endpoint fit is sensitive to arm-order projection, and a regular-design Gaussian endpoint approximation shows mild upper-tail undercoverage without testing the finite-sample outer certificate. EXP8 reports finite-pilot audit rankings that vary by budget and baseline; it does not validate the growing-consistent-pilot allocation guarantee. Protocols and retained outcomes for these diagnostics appear in Section F.3, separately from HelpSteer2. Their released artifacts support result inspection but not end-to-end execution; EXP8 retains historical summaries and only partial raw evidence, as detailed in Section J.7.

The supplementary Open Bandit analysis retains real fixed-action reward/propensity geometry with injected mismatch. Its empirical intervals and audit quantiles assess containment of the randomized-log mean benchmark. The population-valid certificate in Theorem D.3 addresses the separate generated-pair target.

K.3 Integrity and interpretation checks

The frozen formal matrix contains $25{,}000$ completed, zero failed, and zero pending rows in `HelpSteer2/results/formal/raw_results.parquet`. Recomputing the summary from that table verifies the condition denominators, inclusion links, widths, and point biases. The primary complete-chain event is stronger than scalar-truth containment, and every stored chain equals the literal conjunction on these rows. The displayed width decomposition includes a signed empirical residual that reflects training and evaluation variability. Nonvacuity excludes some legal effects, but all final intervals include zero.

The historical freeze manifest records the pre-run protocol, code, and full prepared-data hashes. The anonymous release keeps that manifest as provenance and separately validates the deidentified runtime projection, current release files, and saved numerical results.

The empirical results describe the fixed corpus, discrete rating marginals, constructed selector family, and prescribed controls. Further studies could examine a second corpus, naturally recorded selector

evolution, repeated human counterfactual feedback, and new historical audits. All frozen conditions are retained.