Mechanism Interferometry

A Gauge-Invariant Soft-Intervention Certificate for Causal Modularity

Jeffrey Emanuel

@Dicklesworthstone

August 2026

Abstract

We study a system observed under a reference regime, primitive perturbations, and combinations of those perturbations. For common-support soft interventions, autonomous causal mechanisms impose a rigid algebra on the family of regime distributions. A primitive intervention that replaces one conditional mechanism has a likelihood ratio local to the target and its parents. The ratio conditionally normalizes over the target. Distinct mechanism replacements compose multiplicatively, so mixed finite differences of regime log densities vanish on every square of the factorial design.

These facts yield a finite biconditional certificate on a complete factorial cube (or complete subcube). Relative to a proposed directed acyclic graph and target assignment, locality, targetwise conditional normalization, and zero square curvature on every face are necessary and sufficient for the existence of a context-invariant modular soft-intervention representation. On an irregular partial design, observed squares suffice only when their contrasts span the full lack-of-fit space; in general the certificate instead tests every contrast in the null space of the observed main-effects design. A master identity,

$$\mathbb{E}_0[r_e(X) \mid X_A] = \frac{p_e(X_A)}{p_0(X_A)},$$

turns orientation into ordinary marginal invariance testing: within a localized causal family, the target is the coordinate whose deletion leaves the remaining marginal unchanged. The number of invariant deletions is itself diagnostic: undercoverage degrades gracefully, affected descendants generically produce no pass, and balanced parity mechanisms can produce multiple passes.

For two perturbations we define the gauge-invariant mechanism-curvature field

$$\kappa_{AB}(x) = \log \frac{p_{AB}(x)p_0(x)}{p_A(x)p_B(x)}.$$

It is the conditional design log-odds ratio after subtracting the known sampling-design odds. Full-state modularity implies $\kappa \equiv 0$; the converse additionally requires the certificate's locality and conditional-normalization conditions. Marginalization can create curvature even when the complete mechanisms compose flatly. We derive the exact identity

$$e^{\kappa_{AB}(x)} - 1 = \frac{\mathrm{Cov}_0(L_A, L_B \mid X = x)}{\mathbb{E}_0[L_A \mid X = x]\mathbb{E}_0[L_B \mid X = x]},$$

its Louis missing-information limit, three conservation laws connecting latent conditional coupling to observable ratio-field anticorrelation, and a nested state-expansion identity that turns curvature into a search criterion for missing measurements.

The inferential architecture has two complementary tracks. With arbitrary regime-conditioned samples and known, state-independent sampling proportions, four-law functionals M_w identify weighted curvature without an assignment model. Under product factorial randomization, projected curvature is tested by the Generalised Covariance Measure and its weighted extension, applied to the design bits. Localization uses the same residual-product machinery; orientation uses simultaneous low-dimensional equivalence tests; flexible multinomial models provide diagnostic interaction fields and compositional prediction. A running example exhibits maximal outcome synergy with zero complete-state curvature and nonzero outcome-only curvature in closed form. Simulations cover symmetry-induced orientation failure, latent-state conservation, and implementation inconsistency. We conclude with protocols for feature-flag experiments, combinatorial Perturb-seq, representation learning, and a memory-safe Rust implementation built on the FrankenPandas, FrankenNumPy, FrankenSciPy, and FrankenTorch ecosystems.

Keywords: causal modularity; soft interventions; density ratios; conditional odds ratios; factorial experiments; mechanism changes; causal representation learning; missing information; conditional independence.

Contents

1 Introduction **5**

1.1 Contributions . . . 6

2 Setup and scope **7**

2.1 Baseline causal model . . . 7

2.2 Modular soft interventions . . . 7

3 Locality, normalization, and the augmented regime graph **8**

3.1 A global perturbation has a local ratio . . . 8

3.2 Conditional normalization orients the family . . . 8

4 The gauge field on the intervention cube **9**

4.1 Mixed finite differences . . . 9

4.2 Gauge invariance . . . 9

4.3 Conditional design odds . . . 9

5 The finite certificate **10**

6 The master marginal identity and target orientation **11**

7 Support errors, pass counts, and graph recovery **12**
7.1 Undercoverage . . . 12
7.2 Overcoverage by unaffected nondescendants . . . 13
7.3 Affected descendants . . . 13
7.4 Pass-count audit . . . 13
7.5 A nonpathological parity failure . . . 13
7.6 Active tilt selection . . . 14
7.7 Recovering the DAG from family sets . . . 15

8 Conservation laws **15**
8.1 The moment battery . . . 15
8.2 Latent-factor conservation . . . 16

9 Running example: synergy, flatness, and marginal curvature **17**
9.1 Complete state . . . 17
9.2 Outcome synergy . . . 17
9.3 Retaining only the outcome . . . 17

10 Infinitesimal curvature and missing Fisher information **19**

11 Nested state expansion **19**

12 Statistical inference **20**
12.1 Track I: functionals of the four regime laws . . . 20
12.2 Track II: randomized product factorial designs . . . 21
12.3 The product-sampling compatibility condition . . . 21
12.4 Joint multinomial estimation . . . 22
12.5 Self-normalized compositional prediction . . . 23
12.6 Localization by Design-IAMB . . . 23
12.7 Orientation requires equivalence inference . . . 24
12.8 Overlap and selection gates . . . 24
12.9 Estimator-family agreement . . . 24
12.10 One statistical toolbox . . . 25

13 Partial factorial designs **25**

14 Representation learning **26**

15 Simulation suite **27**
15.1 Complete versus marginalized running example . . . 27
15.2 Parity multi-pass failure . . . 27
15.3 Latent conservation . . . 27

15.4 Implementation inconsistency . . . 28

16 Empirical programs **28**

16.1 Engineered feature flags . . . 28

16.2 Combinatorial Perturb-seq . . . 29

17 Failure taxonomy and limitations **30**

18 Relation to prior work and novelty **30**

19 Conclusion **32**

A Additional proofs **32**

A.1 Square flatness implies all-order flatness . . . 32

A.2 Peeling reconstruction from unlabeled families . . . 32

A.3 Preservation under design sufficiency . . . 33

A.4 Nested state expansion . . . 33

B Partial-design estimability **33**

C Simulation derivations **34**

C.1 Gaussian ratio identity . . . 34

C.2 Parity arithmetic . . . 34

C.3 Implementation factor normalization . . . 34

D Longitudinal extension **34**

E Software audit contract **34**

1 Introduction

Empirical interaction is easy to detect and hard to interpret. A pair of perturbations can look synergistic because the response function is nonlinear, because the perturbations alter one another's mechanisms, because the measured state omits variables that would separate the mechanisms, or because the combined intervention is implemented differently from the two single interventions. Standard outcome-scale interaction statistics generally mix these possibilities. A causal analysis needs an object that lives at the level of the complete data-generating law rather than at the level of one selected response summary.

This paper develops such an object from a simple cancellation argument. Suppose a baseline density factorizes over a directed acyclic graph (DAG), and a soft intervention replaces exactly one conditional mechanism. In the likelihood ratio between the perturbed and baseline regimes, every unchanged mechanism cancels. The ratio is therefore supported only on the target and its direct parents. Because the replacement is a normalized conditional law, the ratio has conditional mean one over the target. When several distinct replacements are applied together, their ratios multiply. In log-density space, autonomous mechanism changes add exactly even when observable outcomes interact violently.

The resulting calculus is most transparent when the complete factorial regime family is treated as a single function-valued log-linear model,

$$\log p_s(x) = \log p_0(x) + \sum_{j=1}^{K} s_j \ell_j(x), \qquad s \in \{0,1\}^K. \tag{1}$$

The absence of design-by-design-by-state interactions is equivalent to zero mixed finite differences of $\log p_s(x)$ on the design cube. The potentials ℓ_j depend on a reference corner; their mixed differences do not. This motivates the name *mechanism interferometry*: the experiment is read not only through the displacement of each regime, but through cycle-closure residuals among regimes.

The closest ingredients are classical or already present in neighboring literatures. Local mechanism shifts underlie causal discovery from changes, uncertain intervention targets, and Joint Causal Inference (Tian and Pearl, 2001; Eaton and Murphy, 2007; Mooij et al., 2020; Squires et al., 2020; Brouillard et al., 2020). Radon–Nikodym ratios are ordinary change-of-measure objects, recently emphasized as causal density functions (Mahadevan, 2026a). The pairwise product law and four-corner contrast have been used directly to discover interactions from unstructured data (Xu et al., 2024). Conditional odds ratios, collapsibility, missing-information decompositions, and residual-product conditional-independence tests have mature theories (Chen, 2007; Tchetgen Tchetgen et al., 2010; Greenland et al., 1999; Louis, 1982; Shah and Peters, 2020; Scheidegger et al., 2022). The contribution here is the assembly: these facts are made to constrain one another in an overdetermined certificate and an executable audit protocol.

1.1 Contributions

The paper makes eight main contributions.

1. **A finite biconditional certificate.** For a complete factorial family of common-support soft interventions with proposed distinct targets, locality, conditional normalization, and zero curvature on every two-dimensional design face are necessary and sufficient for a modular causal representation. Pairwise square flatness then suffices and all higher Boolean-lattice curvatures vanish automatically; irregular partial designs require the full estimable lack-of-fit contrast space.
2. **Ratio-free orientation.** The identity $\mathbb{E}_0[r_e \mid X_A] = p_e(X_A)/p_0(X_A)$ converts target orientation into low-dimensional two-sample equivalence tests. It also pins the otherwise unidentified additive constant of an unnormalized discriminative log-ratio score.
3. **Support-error diagnostics.** The deletion rule is robust to omitted parents and unaffected nondescendants, but affected descendants generically make every deletion fail. Zero, one, or multiple invariant deletions become model diagnostics rather than merely estimator outputs.
4. **Gauge-invariant curvature and conservation laws.** We identify curvature with excess conditional design log odds and derive exact balance identities linking weighted curvature, observable ratio covariance, and hidden conditional likelihood-factor covariance.
5. **A missing-information interpretation.** Infinitesimal observed curvature is the conditional covariance of complete-state intervention scores, precisely the off-diagonal missing-information term in Louis's decomposition. A nested covariance identity shows how adding measurements recovers missing mechanism information in stages.
6. **Two coherent inference tracks.** Four-law moment functionals apply to regime-conditioned samples with arbitrary known state-independent sampling proportions. Under product factorial randomization, the GCM and weighted GCM provide orthogonal, cross-fitted tests of projected curvature. The distinction prevents a subtle but consequential misuse of residual-product tests under arbitrary corner quotas.
7. **Partial-design geometry.** On an arbitrary observed design, testable flatness is exactly the lack-of-fit space of a main-effects-only design matrix with a function-valued response. Classical fractional-factorial aliasing theory therefore applies pointwise in state space.
8. **A complete empirical and software protocol.** We specify localization, orientation, curvature testing, overlap gates, compositional prediction, state expansion, simulations, engineered feature-flag validation, Perturb-seq analysis, and a safe-Rust implementation architecture.

2 Setup and scope

2.1 Baseline causal model

Let $X = (X_1, \ldots, X_d)$ be a complete causal state taking values in a standard measurable space. Under a baseline regime, assume the density p_0 factorizes over a DAG G:

$$p_0(x) = \prod_{i=1}^{d} p_i^0(x_i \mid x_{\mathrm{pa}(i)}). \tag{2}$$

The reference measure is suppressed. Equalities involving densities hold almost everywhere on their common support.

There are K primitive perturbations. A design corner is $s = (s_1, \ldots, s_K) \in \{0,1\}^K$, and P_s denotes the state law under that corner. The all-zero corner is denoted 0. The primitive corner that activates only perturbation j is e_j.

Assumption 2.1 (Common support)**.** *For every regime used in a ratio or cycle contrast, P_s and P_0 are mutually absolutely continuous on a common support.*

Define the density ratio and log-density ratio

$$r_s(x) = \frac{p_s(x)}{p_0(x)}, \qquad \ell_s(x) = \log r_s(x). \tag{3}$$

Common support is not cosmetic. Under P_0,

$$\mathrm{Var}_0(r_s) = \mathbb{E}_0[r_s^2] - 1 = \chi^2(P_s \| P_0), \tag{4}$$

so power and estimator stability deteriorate directly with overlap. Deterministic hard interventions on continuously distributed variables are often singular; the present theory is therefore a soft-intervention calculus. Infinitesimal score fields are the natural limit object when perturbations become small.

2.2 Modular soft interventions

Definition 2.2 (Primitive modular replacement)**.** *Primitive perturbation j targets node t_j if it replaces the baseline conditional mechanism $p_{t_j}^0$ by a valid conditional density q_j, while all other conditional mechanisms remain unchanged:*

$$p_{e_j}(x) = q_j(x_{t_j} \mid x_{\mathrm{pa}(t_j)}) \prod_{i \neq t_j} p_i^0(x_i \mid x_{\mathrm{pa}(i)}). \tag{5}$$

Definition 2.3 (Context-invariant composition)**.** *Primitive replacements compose context-invariantly if, whenever perturbation j is active in a combination, it uses the same q_j that it uses alone. For* *distinct targets,*

$$p_s(x) = \prod_{j:s_j=1} q_j(x_{t_j} \mid x_{\mathrm{pa}(t_j)}) \prod_{i \notin \{t_j : s_j = 1\}} p_i^0(x_i \mid x_{\mathrm{pa}(i)}). \tag{6}$$

Consistent composition is the most load-bearing empirical assumption in the paper. In combinatorial gene perturbation, guide competition, position effects, and dosage changes can make a nominal perturbation different in a pair than alone. In software systems, two feature flags can compete for a shared queue, memory budget, or rate limit. Such effects are scientifically important, but they are implementation curvature rather than evidence that an otherwise correctly implemented complete causal state is insufficient.

3 Locality, normalization, and the augmented regime graph

3.1 A global perturbation has a local ratio

For a primitive replacement at t, all unchanged factors cancel:

$$r_e(x) = \frac{p_e(x)}{p_0(x)} = \frac{q_e(x_t \mid x_{\mathrm{pa}(t)})}{p_t^0(x_t \mid x_{\mathrm{pa}(t)})}. \tag{7}$$

Thus r_e is measurable with respect to

$$F_t = \{t\} \cup \mathrm{pa}(t). \tag{8}$$

Descendants can move dramatically in marginal distribution, but their downstream mechanisms cancel in the joint ratio.

Let E be a binary label distinguishing regime e from baseline, with pooled sampling probabilities ρ_e, ρ_0. Bayes' rule gives

$$\log \frac{\mathbb{P}(E = e \mid X = x)}{\mathbb{P}(E = 0 \mid X = x)} = \log \frac{\rho_e}{\rho_0} + \ell_e(x). \tag{9}$$

Therefore

$$E \perp X_{-F_t} \mid X_{F_t}. \tag{10}$$

In the augmented DAG with context edge $E \to X_t$, F_t is the Markov blanket of E: its child and the child's other parents. Localization is therefore not an escape from conditional-independence testing. It relocates the search from state–state relations to design–state relations, where the target variable is discrete, externally assigned in experiments, and low-dimensional.

3.2 Conditional normalization orients the family

Because $q_e(\cdot \mid x_{\mathrm{pa}(t)})$ is a conditional density,

$$\mathbb{E}_0[r_e(X) \mid X_{\mathrm{pa}(t)}] = 1. \tag{11}$$

This asymmetry distinguishes the child from its parents. The same condition will later be turned into a ratio-free marginal test.

4 The gauge field on the intervention cube

4.1 Mixed finite differences

Let

$$h_s(x) = \log p_s(x). \tag{12}$$

For a set $T \subseteq [K]$ and a base corner b with $b_j = 0$ for every $j \in T$, define

$$\kappa_T^b(x) = \Delta_T h(b; x) = \sum_{U \subseteq T} (-1)^{|T|-|U|} h_{b+\mathbf{1}_U}(x). \tag{13}$$

For two perturbations A, B at the reference corner,

$$\boxed{\kappa_{AB}(x) = \log \frac{p_{AB}(x)p_0(x)}{p_A(x)p_B(x)}.} \tag{14}$$

Equivalently,

$$r_{AB}(x) = r_A(x)r_B(x)e^{\kappa_{AB}(x)}. \tag{15}$$

4.2 Gauge invariance

Adding any field of the form

$$a(x) + \sum_j s_j b_j(x) \tag{16}$$

to $h_s(x)$ leaves every mixed difference of order at least two unchanged. The primitive potentials are reference-dependent, while the curvature is not. Reversing a coordinate orientation changes the sign of any mixed difference containing that coordinate, but the flatness statement $\kappa_T = 0$ is invariant. No scientifically privileged baseline is required.

The analogy with a gauge field is exact at the algebraic level: edge increments are potentials, cycle sums are field strengths, and zero cycle sums are discrete integrability conditions. In free-energy perturbation, analogous cycle-closure residuals diagnose inconsistency among estimated log free-energy differences, and joint multistate estimators pool evidence across states (Shirts and Chodera, 2008). Here the response is not a scalar free energy but the full function $x \mapsto \log p_s(x)$.

4.3 Conditional design odds

Let $A, B \in \{0, 1\}$ be the design bits, and let ρ_{ab} be the state-independent probabilities with which observations from the four regime laws are pooled. Write

$$q_{ab}(x) = \mathbb{P}_\rho(A = a, B = b \mid X = x). \tag{17}$$

Bayes' rule yields

$$\boxed{\kappa_{AB}(x) = \log \frac{q_{11}(x)q_{00}(x)}{q_{10}(x)q_{01}(x)} - \log \frac{\rho_{11}\rho_{00}}{\rho_{10}\rho_{01}}.} \tag{18}$$

Thus curvature is excess conditional design log odds after subtracting the known pooled-design odds. The density functional includes its constant component. A free-intercept multinomial parametrization can hide that component only if the analyst reports a centered interaction term rather than reconstructing the complete posterior odds and subtracting $\log \mathrm{OR}(\rho)$.

Equation (18) survives arbitrary known state-independent corner quotas. It fails under within-regime state-dependent selection: if inclusion depends on X given the regime, the sampled conditional law is not p_s, and the four-law density curvature is no longer recovered without modeling selection.

5 The finite certificate

Theorem 5.1 (Finite modular soft-intervention certificate)**.** *Assume that the complete factorial family $\{p_s : s \in \{0,1\}^K\}$ is observed, p_0 factorizes over a proposed DAG G, all regimes have common support, and primitive perturbations are assigned distinct proposed targets $t_1, \ldots, t_K$. Let*

$$r_j(x) = \frac{p_{e_j}(x)}{p_0(x)}. \tag{19}$$

The following are equivalent.

(A) There exist valid conditional mechanisms $q_j(x_{t_j} \mid x_{\mathrm{pa}(t_j)})$ such that perturbation j replaces only mechanism t_j, the replacement is context-invariant, and the resulting modular factorization reproduces every p_s.

(B) The following three conditions hold:

(i) Locality:

$$r_j(x) = r_j(x_{t_j}, x_{\mathrm{pa}(t_j)}). \tag{20}$$

(ii) Conditional normalization:

$$\mathbb{E}_0[r_j(X) \mid X_{\mathrm{pa}(t_j)}] = 1. \tag{21}$$

(iii) Square flatness: *for every pair $j \neq k$ and every base corner b on which the corresponding two-dimensional face is defined,*

$$\kappa^b_{jk}(x) = 0. \tag{22}$$

Proof. If (A) holds, factor cancellation gives locality, integration of each replacement conditional gives normalization, and context-invariant composition gives

$$\frac{p_s(x)}{p_0(x)} = \prod_j r_j(x)^{s_j}. \tag{23}$$

Every mixed finite difference of order at least two therefore vanishes.

Conversely, define

$$q_j(x_{t_j} \mid x_{\mathrm{pa}(t_j)}) = r_j(x_{t_j}, x_{\mathrm{pa}(t_j)}) p^0_{t_j}(x_{t_j} \mid x_{\mathrm{pa}(t_j)}). \tag{24}$$

Locality gives the correct arguments and normalization makes q_j a valid conditional density. It remains to show that square flatness implies global multiplicative composition. For $s_j = 0$, set

$$\delta_j(s; x) = h_{s+e_j}(x) - h_s(x). \tag{25}$$

Zero (j,k)-face curvature gives

$$\delta_j(s + e_k; x) = \delta_j(s; x). \tag{26}$$

The $(K-1)$-cube with coordinate j fixed is connected, so $\delta_j(s;x)$ is independent of every other design coordinate. Hence it equals its value at the reference corner, $\ell_j(x)$. Summing edge increments along any path from 0 to s,

$$h_s(x) = h_0(x) + \sum_j s_j \ell_j(x). \tag{27}$$

Therefore $p_s = p_0 \prod_j r_j^{s_j}$, which is exactly the density generated by replacing the baseline kernels with (24). This proves (A). □

Corollary 5.2 (Order two suffices)**.** *On a complete Boolean cube or complete subcube, zero curvature on every square implies that every higher-order Möbius coefficient vanishes.*

The theorem is an existence result. It says that a proposed graph and target assignment admit a modular causal realization of the observed regime family. Faithfulness and intervention coverage are required to infer that proposal uniquely from data.

6 The master marginal identity and target orientation

Lemma 6.1 (Master marginal identity)**.** *For every regime e and coordinate subset A,*

$$\boxed{\mathbb{E}_0[r_e(X) \mid X_A] = \frac{p_e(X_A)}{p_0(X_A)}.} \tag{28}$$

Proof. For every bounded measurable $g(X_A)$,

$$\mathbb{E}_0[g(X_A) r_e(X)] = \mathbb{E}_e[g(X_A)] = \mathbb{E}_0\left[g(X_A)\frac{p_e(X_A)}{p_0(X_A)}\right]. \tag{29}$$

Uniqueness of conditional expectation proves the result. □

The unconditional special case is $\mathbb{E}_0[r_e] = 1$. If a discriminative model learns an unnormalized score $g_e(x) = \ell_e(x) + c$, the constant can be pinned on held-out baseline data by

$$\widehat{\ell}_e(x) = g_e(x) - \log \widehat{\mathbb{E}}_0[e^{g_e(X)}]. \tag{30}$$

More importantly, the orientation condition becomes a plain marginal invariance statement. Let S be a candidate support containing the target t. For every $v \in S$, define the deletion hypothesis

$$H_v : \quad P_e^{X_{S\setminus\{v\}}} = P_0^{X_{S\setminus\{v\}}}. \tag{31}$$

By theorem 6.1, this is equivalent to

$$\mathbb{E}_0[r_e(X) \mid X_{S\setminus\{v\}}] = 1. \tag{32}$$

If $S = \{t\} \cup S_P$ with $S_P \subseteq \mathrm{pa}(t)$, deleting t leaves only pre-intervention nondescendants, so H_t holds. Deleting a parent leaves the perturbed target in the tested marginal and generally fails.

Assumption 6.2 (Deletion faithfulness)**.** *For the chosen collection of mechanism tilts, the target is the unique coordinate in the localized support whose deletion leaves an invariant marginal.*

Under deletion faithfulness,

$$\boxed{t = exttheuniquev \in S \text{ such that } P_e^{X_{S\setminus\{v\}}} = P_0^{X_{S\setminus\{v\}}}.} \tag{33}$$

This should be estimated by low-dimensional two-sample equivalence tests, not by conditional means of an estimated density ratio. A global calibration error in $\widehat{r}_e$ moves every estimated conditional mean away from one; direct marginal comparison avoids that failure.

7 Support errors, pass counts, and graph recovery

Let

$$\mathrm{Pass}(\widehat{S}) = \left\{ v \in \widehat{S} : P_e^{X_{\widehat{S}\setminus\{v\}}} = P_0^{X_{\widehat{S}\setminus\{v\}}} \right\}. \tag{34}$$

Assume initially that $t \in \widehat{S}$.

7.1 Undercoverage

If

$$\widehat{S} \subseteq \{t\} \cup \mathrm{pa}(t), \tag{35}$$

but the intervention ratio happens not to depend detectably on one or more true parents, then deleting t still leaves a subset of parents. The target passes, and under deletion faithfulness it remains the unique pass. The recovered parent set is $\widehat{S} \setminus \{t\} \subseteq \mathrm{pa}(t)$. Undercoverage therefore misses edges rather than reversing the target.

7.2 Overcoverage by unaffected nondescendants

If $\widehat{S}$ includes an unaffected nondescendant Z, deleting t still leaves only variables whose joint marginal is invariant. The target can remain the unique passing coordinate. Orientation is robust, but the unpruned candidate parent set can falsely retain Z. Minimal Markov-boundary localization or a backward pruning stage is still required for exact graph recovery.

7.3 Affected descendants

If $\widehat{S}$ contains a descendant D whose marginal law changes, deleting the target leaves D, so the target fails. Deleting any other coordinate also leaves D. Deleting D usually leaves the shifted target family. Consequently, descendant contamination generically produces no passes. "Generically" is essential: with partial family coverage, deleting the contaminant can leave a specially balanced parity-invariant submarginal.

7.4 Pass-count audit

The intended interpretation is therefore

Pass count	Interpretation
Exactly one	Coherent single-target family with identifiable orientation.
Zero	Descendant contamination, multi-target primitive, selection, implementation mismatch, or insufficient support.
More than one	Symmetry, redundancy, deletion-faithfulness failure, or an insufficiently rich set of tilts.

7.5 A nonpathological parity failure

Let $P \sim \text{Bernoulli}(1/2)$, $N \sim \text{Bernoulli}(\varepsilon)$, and define the baseline child

$$T = P \oplus N. \tag{36}$$

Under intervention, replace the mechanism by

$$T = \neg P \oplus N. \tag{37}$$

The ratio depends on (P, T), but both one-dimensional marginals remain balanced. Deleting T leaves the invariant parent marginal; deleting P leaves the invariant child marginal. Both coordinates pass. Balanced parity mechanisms are a live discrete failure mode, not a measure-zero analytic curiosity.

Multiple asymmetric tilts of the same target are the natural remedy. The target must pass every valid tilt; a spurious symmetry usually does not. When the experimental record identifies which tilts replace the same target mechanism, taking the union of their localized supports can also reveal parents omitted by a particular intervention. That grouping is design metadata, not something an unlabeled family multiset generally identifies on its own.

7.6 Active tilt selection

The pass-count taxonomy is not only a verdict; it prescribes the next experiment. Suppose a family reaches the multiple-pass state with competing invariant deletions $v_1, \ldots, v_m$. Each competitor v_i corresponds to the hypothesis that v_i is the target, and each hypothesis predicts, for any candidate tilt q, which deletion marginals would remain invariant under that tilt. A discriminating tilt is one whose predicted deletion outcomes differ across the surviving hypotheses. Given a feasible tilt class $\mathcal{Q}$, select

$$q^\star = \arg\max_{q \in \mathcal{Q}} \min_{i \neq j} D\Big(P_i^{(q)}, P_j^{(q)}\Big), \tag{38}$$

where $P_i^{(q)}$ denotes the deletion-marginal pattern predicted under hypothesis i and D is the same discrepancy used for equivalence testing. Maximizing the worst-case separation targets the residual symmetry directly rather than spending samples uniformly.

In the parity example of section 7.5, the ambiguity is sustained by the symmetry of the replacement: both competing deletion marginals stay balanced. The admissible next experiment is another tilt of the same target mechanism, not an intervention on a different node. Care is needed in choosing it: any replacement of the form $T = \neg P \oplus N'$ with independent noise N', however biased, leaves T exactly uniform, so biasing the noise alone cannot break the tie. An explicitly asymmetric conditional does: take $q(T = 1 \mid P = 0) = \alpha$ and $q(T = 1 \mid P = 1) = \beta$ with $\alpha + \beta \neq 1$ and common support, so that $\mathbb{P}(T = 1) = (\alpha + \beta)/2 \neq 1/2$. The target's own marginal then moves, the parent-deletion marginal changes while the target-deletion marginal remains invariant, and one additional regime separates the two hypotheses. The same criterion covers the state-expansion loop: when curvature is found and candidate measurements compete to explain it, equation (71) predicts each candidate's residual field, and the next measurement is the one whose predictions best separate the surviving explanations. Orientation, like state expansion, thereby becomes a sequential design problem with an explicit objective, rather than a fixed battery that either succeeds or abstains.

The acquisition score in (38) has proposal authority only. It depends on a predictive model for experiments not yet run, so model misspecification can make the ranking poor even when the eventual certificate remains valid. Before collecting follow-up data, freeze the surviving hypotheses, feasible tilt library, predicted pairwise separations, model and data fingerprints, random seed, and deterministic tie rule; the software artifact binds the complete ordered candidate library with a derived content fingerprint. Reject candidates that change a different primitive, lack measurable delivery or common support, or would invoke product-factorial inference without verified product odds or an explicit reweighting plan. The selected tilt is then a new independently randomized regime and re-enters the audit at preflight; neither its predicted separation nor its rank resolves the original multiple-pass state. If no candidate survives these feasibility gates, the correct output is an explicit experiment-design abstention rather than a forced target.

7.7 Recovering the DAG from family sets

If each node receives a sufficiently rich single-target perturbation, orientation gives every family $\{i\} \cup \mathrm{pa}(i)$ directly. Even without target labels, the multiset of family sets uniquely determines a DAG. Process families in topological order. After removing previously identified nodes, the earliest unidentified node has no unidentified parents, so its reduced family is the singleton containing itself. Assign it, restore its removed parents, and continue. A formal proof appears in section A.

The one-family-per-node condition is load bearing. With repeated tilts and no same-target group labels, an undercovered family can peel before a richer family of the same target. For example, the unlabeled multiset $\{\{A\}, \{B\}, \{A, B\}\}$ is compatible with $A \to B$, one source-family observation for A, and two tilts of B, one of which omits parent A. The two singleton sets peel simultaneously, after which the remaining set contains no unidentified node and cannot be assigned uniquely. Thus repeated tilts should be grouped by an independently recorded mechanism/target identifier and unioned before unlabeled peeling; absent that metadata, the reconstruction must abstain rather than infer which already identified member generated the extra family.

8 Conservation laws

8.1 The moment battery

Exact modularity implies

$$\boxed{\mathbb{E}_0\left[\prod_{j\in T} r_j(X)\right] = 1 \quad \text{for every } T \subseteq [K].} \tag{39}$$

For a pair,

$$\mathrm{Cov}_0(r_A, r_B) = 0. \tag{40}$$

The causal families can be statistically dependent; the ratios nevertheless satisfy this orthogonality because their conditional normalizations telescope.

These are inexpensive necessary conditions, but they are not complete. The exact meaning of the pair statistic follows from (15) and $\mathbb{E}_0[r_{AB}] = 1$:

$$\boxed{\mathbb{E}_0[r_A r_B] - 1 = -\mathbb{E}_0\left[r_A r_B(e^{\kappa_{AB}} - 1)\right].} \tag{41}$$

The scalar moment test estimates a signed average of $e^{\kappa}-1$ under the ratio-tilted measure proportional to $r_A r_B\, dP_0$. Its blind spot is quantitative: sign-varying curvature can integrate to zero.

Suppose $r_A r_B \in L^1(P_0)$. For any bounded witness w, define

$$M_w = \mathbb{E}_0\left[w(X)\{r_{AB}(X) - r_A(X) r_B(X)\}\right]. \tag{42}$$

Then

$$M_w = \mathbb{E}_0\left[w(X)r_A(X)r_B(X)(e^{\kappa(X)} - 1)\right]. \tag{43}$$

Under positivity,

$$M_w = 0 \text{ for every bounded } w \quad \Longleftrightarrow \quad \kappa = 0 \quad P_0\text{-a.s.} \tag{44}$$

The scalar battery is the special case $w \equiv 1$.

8.2 Latent-factor conservation

Let the complete state be $Z = (X, U)$, with baseline density q_0. Suppose $L_A, L_B \in L^2(P_0)$ and the complete-state regime factors satisfy

$$q_A = q_0 L_A, \qquad q_B = q_0 L_B, \qquad q_{AB} = q_0 L_A L_B. \tag{45}$$

After marginalizing U,

$$r_A(X) = \mathbb{E}_0[L_A \mid X], \qquad r_B(X) = \mathbb{E}_0[L_B \mid X], \qquad r_{AB}(X) = \mathbb{E}_0[L_A L_B \mid X]. \tag{46}$$

Therefore

$$\boxed{e^{\kappa_{AB}(x)} - 1 = \frac{\mathrm{Cov}_0(L_A, L_B \mid X = x)}{r_A(x) r_B(x)}.} \tag{47}$$

Because q_A, q_B, q_{AB} are normalized,

$$\mathbb{E}_0[L_A] = \mathbb{E}_0[L_B] = \mathbb{E}_0[L_A L_B] = 1, \tag{48}$$

so $\mathrm{Cov}_0(L_A, L_B) = 0$. The law of total covariance gives

$$\boxed{\mathrm{Cov}_0(r_A, r_B) = -\mathbb{E}_0[\mathrm{Cov}_0(L_A, L_B \mid X)].} \tag{49}$$

Combining (47) and (49) yields the three-way conservation law

$$\boxed{\mathrm{Cov}_0(r_A, r_B) = -\mathbb{E}_0[r_A r_B (e^{\kappa} - 1)] = -\mathbb{E}_0[\mathrm{Cov}_0(L_A, L_B \mid X)].} \tag{50}$$

Whatever conditional coupling marginalization creates between flat latent factors must reappear, with opposite sign on average, as correlation between the observable primitive ratio fields. At orders three and above, Brillinger's conditioning formula decomposes cumulants through sums over set partitions; the clean two-term split in (49) is special to covariance (Brillinger, 1969).

9 Running example: synergy, flatness, and marginal curvature

Let $X_1, X_2 \in \{-1, +1\}$ be independent Rademacher variables under baseline, and let

$$Y = X_1 X_2 + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2). \tag{51}$$

Perturbation A tilts the source distribution of X_1:

$$P_A(X_1 = x) = \frac{1 + ax}{2}, \qquad |a| < 1. \tag{52}$$

Perturbation B analogously tilts X_2 by b.

9.1 Complete state

The primitive ratios are

$$r_A(X_1) = 1 + aX_1, \qquad r_B(X_2) = 1 + bX_2. \tag{53}$$

The combined ratio is their product, so

$$\kappa_{AB}(X_1, X_2, Y) = 0. \tag{54}$$

The source mechanisms compose autonomously.

9.2 Outcome synergy

Yet

$$\mathbb{E}_0[Y] = \mathbb{E}_A[Y] = \mathbb{E}_B[Y] = 0, \qquad \mathbb{E}_{AB}[Y] = ab. \tag{55}$$

Thus the additive outcome interaction is exactly ab. Neither perturbation alone changes the mean; together they do. Outcome interaction is a property of a chosen response and scale, whereas mechanism curvature is a property of the full regime law. This sharpens the distinction between statistical and mechanistic interaction emphasized in classical interaction theory (VanderWeele and Knol, 2014).

9.3 Retaining only the outcome

Under either primitive tilt alone, the product $X_1 X_2$ remains symmetric, so

$$p_A(y) = p_B(y) = p_0(y). \tag{56}$$

Under the combined tilt,

$$p_{AB}(y) = \frac{1 + ab}{2}\phi_\sigma(y - 1) + \frac{1 - ab}{2}\phi_\sigma(y + 1), \tag{57}$$

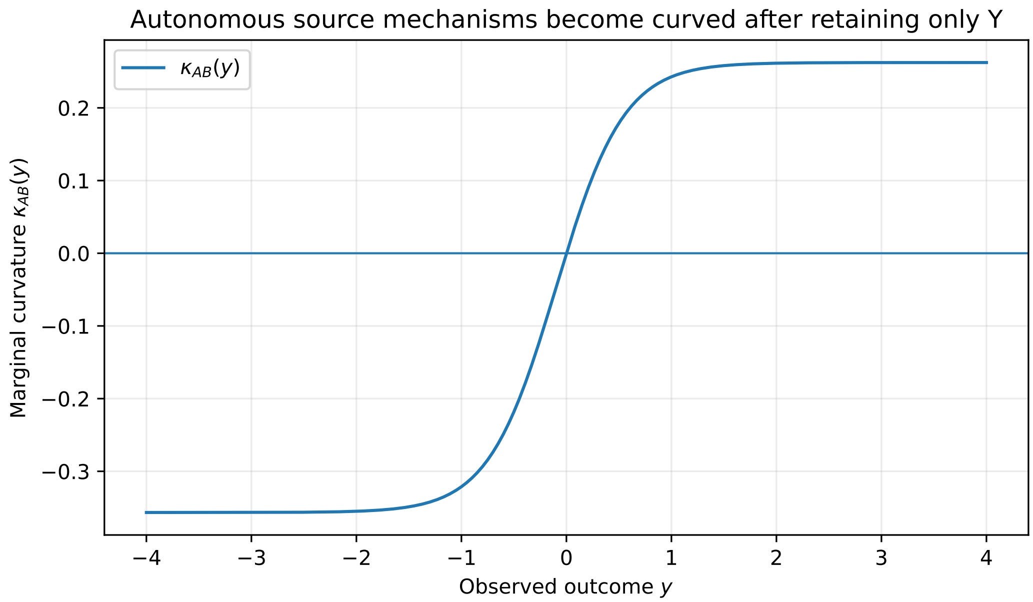

Figure 1: For $a = 0.6$, $b = 0.5$, and $\sigma = 0.8$, the complete source mechanisms are flat while the outcome-only representation has sign-varying curvature.

while

$$p_0(y) = \frac{1}{2}\phi_\sigma(y-1) + \frac{1}{2}\phi_\sigma(y+1). \tag{58}$$

Using

$$\frac{\phi_\sigma(y-1) - \phi_\sigma(y+1)}{\phi_\sigma(y-1) + \phi_\sigma(y+1)} = \tanh\left(\frac{y}{\sigma^2}\right), \tag{59}$$

we obtain

$$\frac{p_{AB}(y)}{p_0(y)} = 1 + ab\tanh\left(\frac{y}{\sigma^2}\right) \tag{60}$$

and therefore

$$\boxed{\kappa_{AB}(y) = \log\left[1 + ab\tanh\left(\frac{y}{\sigma^2}\right)\right].} \tag{61}$$

The example is also the canonical blind spot of the scalar moment battery. In the Y-only representation,

$$r_A(y) = r_B(y) = 1, \tag{62}$$

so $\mathbb{E}_0[r_A r_B] = 1$ passes trivially. Nevertheless $\kappa(y) \not\equiv 0$. The weighted mean vanishes because p_0 is symmetric and $\tanh(y/\sigma^2)$ is odd.

The latent covariance cross-check is immediate:

$$\mathbb{E}_0[X_1 \mid Y = y] = \mathbb{E}_0[X_2 \mid Y = y] = 0, \tag{63}$$

while

$$\mathbb{E}_0[X_1 X_2 \mid Y = y] = \tanh(y/\sigma^2). \tag{64}$$

Therefore

$$\mathrm{Cov}_0(L_A, L_B \mid Y = y) = ab \tanh(y/\sigma^2), \tag{65}$$

which reproduces (61) through (47).

10 Infinitesimal curvature and missing Fisher information

Let $q_{\alpha,\beta}(z)$ be a smooth complete-state intervention family, and assume the usual dominated regularity conditions that justify differentiating twice under the integral and taking the displayed conditional moments. Let

$$p_{\alpha,\beta}(x) = \int q_{\alpha,\beta}(x, u)\, du. \tag{66}$$

Define complete-data scores

$$S_A(z) = \partial_\alpha \log q_{\alpha,\beta}(z)\big|_{0,0}, \qquad S_B(z) = \partial_\beta \log q_{\alpha,\beta}(z)\big|_{0,0}. \tag{67}$$

Differentiating under the integral and evaluating at the reference parameter gives

$$\begin{aligned} \partial^2_{\alpha\beta} \log p_{\alpha,\beta}(x)\Big|_{0,0} &= \mathbb{E}_0\left[\partial^2_{\alpha\beta} \log q_{\alpha,\beta}(Z)\Big|_{0,0} \middle| X = x \right] \\ &\quad + \mathrm{Cov}_0(S_A, S_B \mid X = x). \end{aligned} \tag{68}$$

If the complete-state intervention model has no cross derivative, then

$$\boxed{\mathcal{C}_{AB}(x) := \partial^2_{\alpha\beta} \log p_{\alpha,\beta}(x)\Big|_{0,0} = \mathrm{Cov}_0(S_A, S_B \mid X = x).} \tag{69}$$

This covariance is the off-diagonal missing-information term in Louis's observed-information decomposition (Louis, 1982). With the usual convention that observed information is minus the observed log-likelihood Hessian, curvature is the negative of the corresponding observed-information cross entry when the complete cross-Hessian is zero.

This supplies a power heuristic: infinitesimal curvature is detectable only insofar as the complete intervention scores retain conditional covariance after the observed state is known. It also differentiates the present symmetric mixed derivative from the antisymmetric Lie bracket in infinitesimal intervention geometry (Mahadevan, 2026b). The Lie bracket measures order-sensitive noncommutativity of sequential flows; Boolean curvature measures finite, order-free failure of density-ratio composition.

11 Nested state expansion

Suppose the current observation is X, and a candidate additional measurement is W. Define

$$r_A^{XW} = \mathbb{E}[L_A \mid X, W], \qquad r_B^{XW} = \mathbb{E}[L_B \mid X, W], \tag{70}$$

and define curvature at each observation level through (47). The conditional law of total covariance gives

$$\begin{aligned} & r_A^X r_B^X (e^{\kappa_X} - 1) \\ & \quad = \mathbb{E}\left[r_A^{XW} r_B^{XW} (e^{\kappa_{XW}} - 1) \mid X \right] + \operatorname{Cov}(r_A^{XW}, r_B^{XW} \mid X). \end{aligned} \tag{71}$$

Proposition 11.1 (State-expansion conservation)**.** *Equation* (71) *holds whenever the square-integrable complete-state factors satisfy* (45)*. If W supplies the missing separator so that $\kappa_{XW} = 0$, then*

$$r_A^X r_B^X (e^{\kappa_X} - 1) = \operatorname{Cov}(r_A^{XW}, r_B^{XW} \mid X). \tag{72}$$

A candidate variable explains coarse curvature when the residual interaction field becomes equivalent to zero after adding it and the former curvature is accounted for by variation of the now-separated primitive ratio fields across that variable. This turns curvature from a verdict into a search direction.

Flattening is not monotone. Conditioning on a collider, a selection variable, or a noisy proxy can create curvature. Candidate measurements must therefore be evaluated by nested held-out tests, not greedily accepted because a training loss decreases. Infinitesimally, the same tower property decomposes missing Fisher information across nested sigma-algebras: state expansion recovers missing information in stages.

12 Statistical inference

The population identities are exact; empirical use requires explicit sampling contracts. Two inference tracks cover different data-generating situations.

12.1 Track I: functionals of the four regime laws

Suppose independent samples are available from P_0, P_A, P_B, P_{AB}. Sampling proportions may be arbitrary, provided they are state-independent within each regime. Curvature is a functional of the four normalized laws and is identified independently of a population assignment mechanism.

For a bounded witness w, the moment (42) can be written

$$M_w = \mathbb{E}_{AB}[w(X)] - \mathbb{E}_A[w(X) r_B(X)] \tag{73}$$

or symmetrically with A and B reversed. The first term requires no ratio estimate; the second requires only one primitive ratio nuisance. Cross-fitting, truncation diagnostics, and a symmetrized estimator reduce overfitting and variance. Parametric or semiparametric models for $\kappa(x;\beta)$ can use the conditional odds-ratio machinery of Chen (2007) and Tchetgen Tchetgen et al. (2010).

This track remains valid under non-product corner quotas because the regime laws themselves are the estimands. It fails if the within-regime sample is selected as a function of X and the selection mechanism is ignored.

12.2 Track II: randomized product factorial designs

Now suppose (A, B) are sampled with product probabilities

$$\rho_{ab} = \rho_a^A \rho_b^B. \tag{74}$$

Let

$$m_A(x) = \mathbb{P}(A = 1 \mid X = x), \qquad m_B(x) = \mathbb{P}(B = 1 \mid X = x). \tag{75}$$

For binary variables,

$$\begin{aligned} c(x) &:= \mathrm{Cov}(A, B \mid X = x) && (76)\\ &= q_{11}(x)q_{00}(x) - q_{10}(x)q_{01}(x) && (77)\\ &= q_{10}(x)q_{01}(x)(e^{\kappa(x)} - 1), && (78) \end{aligned}$$

where the last equality uses $\mathrm{OR}(\rho) = 1$. Under positivity,

$$c(x) = 0 \quad \Longleftrightarrow \quad \kappa(x) = 0. \tag{79}$$

For a fixed witness w, define

$$\theta_w = \mathbb{E}[w(X)c(X)]. \tag{80}$$

Then

$$\boxed{\theta_w = \mathbb{E}\left[w(X)(A - m_A(X))(B - m_B(X))\right].} \tag{81}$$

This is the Generalised Covariance Measure of Shah and Peters (2020), applied to the randomized design bits. Weighted collections of witnesses and max-statistic procedures are instances of the weighted GCM of Scheidegger et al. (2022). Cross-fitted nuisance regressions yield Neyman-orthogonal scores; the leading remainder is proportional to the product of the two regression errors, as in double/debiased machine learning (Chernozhukov et al., 2018).

A practical adaptive witness uses sample splitting. Estimate $\widehat{c}(x)$ or $\widehat{\kappa}(x)$ on one fold, set $w(x) = \widehat{c}(x)$ on an independent fold, and test θ_w. Splitting preserves validity while directing local power toward the estimated curvature pattern.

12.3 The product-sampling compatibility condition

The residual-product test is not invariant to arbitrary corner quotas. Under pooled probabilities ρ,

$$\mathrm{Cov}(A, B \mid X = x) = q_{10}(x)q_{01}(x)\left(\mathrm{OR}(\rho)e^{\kappa(x)} - 1\right). \tag{82}$$

Therefore $c(x) \equiv 0$ characterizes

$$\kappa(x) \equiv -\log \mathrm{OR}(\rho), \tag{83}$$

not $\kappa \equiv 0$, unless $\mathrm{OR}(\rho) = 1$. Equal-corner balancing, $\rho_{ab} = 1/4$, is safe when it follows from the sampling contract; equal empirical counts alone do not establish product assignment. Arbitrary per-corner quotas are not safe. The implementation must either verify product sampling from the declared design, reweight observations to a chosen product design, or use Track I. Accordingly, every software GCM projection carries a serialized design-evidence reference: a sampling-odds audit identifier and fingerprint of its declared allocation source, a completed product-design reweighting audit and artifact fingerprint, or an explicit diagnostic-only grade. Identifiers and fingerprints are not self-authenticating; the engine must resolve them against the content-bound manifest, allocation artifact, analyzed data, and evidence ledger before a projection can support certification.

This is the precise correspondence:

	Four-law track	Randomized factorial track
Estimand	M_w, defined solely by P_0, P_A, P_B, P_{AB}	θ_w, a projected conditional design covariance
Sampling	Arbitrary state-independent regime quotas	Product assignment or reweighting to product assignment
Nuisance	Density ratios / odds-ratio model	Conditional design regressions m_A, m_B
Strength	Assignment-free density functional	Orthogonal residual-product score and wGCM theory
Failure	Poor overlap or state-dependent within-regime selection	Non-product quotas silently shift the null

12.4 Joint multinomial estimation

For visualization, flexible diagnostics, and compositional prediction, fit one multinomial model over all observed regime corners. Let

$$\eta_s(x) = \log \frac{\mathbb{P}_\rho(S = s \mid X = x)}{\mathbb{P}_\rho(S = 0 \mid X = x)} - \log \frac{\rho_s}{\rho_0} = \log \frac{p_s(x)}{p_0(x)}. \tag{84}$$

Use a hierarchical design expansion

$$\eta_s(x) = \sum_j s_j f_j(x) + \sum_{j<k} s_j s_k f_{jk}(x) + \cdots . \tag{85}$$

The restricted modular model sets all higher-order fields to zero. The complete fitted posterior odds, including corner intercepts, must be used to reconstruct curvature. A held-out proper-loss advantage of the unrestricted model is a useful diagnostic, but for flexible learners it is not by itself a calibrated hypothesis test.

A joint model pools statistical strength and avoids subtracting independently estimated log ratios with alternating signs. It also makes held-out combination prediction natural.

12.5 Self-normalized compositional prediction

Estimated primitive ratios do not generally multiply to a normalized density. For a held-out combination T, let

$$w_T(x) = \prod_{j \in T} \widehat{r}_j(x), \qquad \widehat{Z}_T = \widehat{\mathbb{E}}_0[w_T(X)]. \tag{86}$$

Use

$$\widetilde{r}_T(x) = \frac{w_T(x)}{\widehat{Z}_T}, \qquad d\widetilde{P}_T = \widetilde{r}_T \, dP_0. \tag{87}$$

Report $\widehat{Z}_T - 1$ rather than silently discarding it: it is itself a failed moment restriction. Compare $\widetilde{P}_T$ with held-out samples from P_T using proper scores, MMD, energy distance, or prespecified transported moments.

12.6 Localization by Design-IAMB

A concrete localization procedure treats the regime label as the response in Markov-boundary discovery. A forward phase repeatedly adds the state variable with the largest reproducible cross-fitted improvement in proper regime-prediction loss, which at the Bayes optimum equals conditional mutual information. A backward phase removes variables whose conditional contribution disappears. This is an intervention-adapted form of IAMB (Tsamardinos et al., 2003).

The conditional-independence engine can again be residual products: one-hot regime residuals against residualized candidate variables. Repeated sample splits and stability selection (Meinshausen and Bühlmann, 2010) are preferable to pretending that a fully nonparametric finite-sample support guarantee is available.

An ensemble alternative to the greedy path exploits the parsimony corollary directly. Sample many random candidate supports, fit a regime classifier on each, and score each support by held-out proper regime-prediction loss. Retain the parsimony frontier: the supports within a fixed tolerance of the best held-out loss, ordered first by support cardinality and only then by a learner-specific complexity measure. Per-variable inclusion frequencies across the frontier are then reported as localization stability paths in the audit bundle. Under ratio faithfulness and adequate learner capacity, locality makes the true support $\{t\} \cup \mathrm{pa}(t)$ the smallest support carrying full regime information, so the frontier filter has a causal justification here that generic model selection lacks; without faithfulness, an inactive true parent can be omitted, which is the familiar undercoverage mode rather than a reversal. Three disciplines keep the procedure honest: the frontier threshold is computed once over the completed ensemble, never incrementally as results accumulate; inclusion frequencies are descriptive frequencies over a designed candidate ensemble, not probabilities; and because the support is selected adaptively, certificate-grade conclusions about the selected support require an outer held-out confirmation sample, with same-sample results labeled exploratory. The deletion pass-count audit then independently checks the localized support.

12.7 Orientation requires equivalence inference

Failure to reject equality is not evidence for invariance. Choose a nonnegative marginal discrepancy D, such as MMD (Gretton et al., 2012) or energy distance (Székely and Rizzo, 2013). Population squared MMD is nonnegative, but its unbiased finite-sample U-statistic can be negative under the null; that signed estimator belongs inside a calibrated resampling procedure, not directly in the nonnegative relative ratio below. A biased MMD V-statistic is one admissible nonnegative point discrepancy. For each deletion,

$$D_v = D(P_e^{X_{S\setminus\{v\}}}, P_0^{X_{S\setminus\{v\}}}), \tag{88}$$

and define the full-support intervention discrepancy

$$D_{\text{full}} = D(P_e^{X_S}, P_0^{X_S}). \tag{89}$$

Use a relative equivalence ratio

$$R_v = \frac{D_v}{D_{\text{full}} + \eta} \tag{90}$$

with a preregistered tolerance ε and small numerical stabilizer η. Relative calibration makes the threshold dimensionless and scales it to intervention strength. A max-statistic bootstrap over all deletions provides simultaneous confidence bounds because the tests share samples and nested marginals.

Classify a deletion as certified invariant if its upper bound is below ε, certified changed if its lower bound exceeds ε, and otherwise undetermined. Orient only when exactly one deletion is certified invariant and every competitor is certified changed. If D_{full} is too small, abstain: an intervention too weak to detect cannot orient a family reliably.

12.8 Overlap and selection gates

Every report should include estimated second moments of ratios, effective sample sizes, tail quantiles and maxima of weights, posterior calibration, sensitivity to truncation, and a declared abstention rule. A classifier can separate regimes perfectly precisely when density-ratio estimation is least reliable. State-dependent missingness, survival, or inclusion must be modeled or treated as a separate selection mechanism; otherwise curvature can be entirely artifactual.

12.9 Estimator-family agreement

Curvature is gauge-invariant in population, but every empirical curvature field, nuisance regression, and multinomial fit inherits the inductive bias of its learner. A verdict that changes when the nuisance family changes is not yet evidence about the system. The audit therefore repeats the primary projections with a small battery of deliberately dissimilar model families, for example a regularized linear model, a kernel or nearest-neighbour method, and a boosted tree or neural learner, each within its own cross-fitting scheme. Certified conclusions require agreement across the battery up to a preregistered tolerance; disagreement is recorded as a finding and, in strict mode, forces abstention. Because the families share data and folds, scaled pairwise gaps are a preregistered sensitivity heuristic rather than a calibrated joint test, and the gate is asymmetric: disagreement blocks, while agreement remains diagnostic and certifies nothing by itself. This is the empirical counterpart of gauge invariance: the population object does not depend on the analyst's parametrization, so the estimate should not depend materially on the analyst's function class. The battery is a robustness gate, not a validity substitute; it cannot repair a violated sampling contract.

12.10 One statistical toolbox

The full pipeline reduces largely to two reusable primitives:

1. weighted, cross-fitted residual products for localization and factorial curvature projections;
2. low-dimensional two-sample equivalence statistics for deletion orientation and held-out distributional prediction.

This unification is operationally important. It replaces a collection of bespoke tests with one residualization engine, one kernel/distance engine, shared cross-fitting, and one simultaneous-resampling layer.

13 Partial factorial designs

Pointwise in x, flatness says that the vector $h_D(x) = (h_s(x) : s \in D)$ lies in the column space of a main-effects design matrix. Let

$$M_D = [\mathbf{1}, s_1, \ldots, s_K]_{s\in D}. \tag{91}$$

The complete testable flatness content on an observed design $D \subseteq \{0,1\}^K$ is

$$C^\top h_D(x) = 0 \qquad \text{for every } C \in \ker(M_D^\top). \tag{92}$$

Thus estimability and aliasing are exactly those of classical factorial design, with a function-valued response (Box et al., 2005; Wu and Hamada, 2009).

Squares suffice when observed face contrasts span the lack-of-fit space, as in complete subcubes and many square-connected designs. They need not suffice on irregular designs. In the three-cube with corners 000 and 111 removed, no square face is fully observed, yet the six-point main-effects matrix has rank four, leaving a two-dimensional lack-of-fit space. One valid contrast is

$$h_{110} - h_{100} - h_{011} + h_{001}, \tag{93}$$

which compares a coordinate increment across bases that differ in two other coordinates.

For engineered systems, this is not merely a caveat. Full 2^K experiments are infeasible for large

K. Resolution-V fractions can make all two-factor interaction fields estimable without aliasing them with mechanism potentials. A complete treatment belongs in a follow-up; section B gives the formal linear-algebra statement and a design-audit algorithm.

14 Representation learning

The desired causal coordinates are those in which regime changes are local, normalized, and flat. But minimizing a flatness penalty directly is gameable: an encoder can discard exactly the coordinates on which curvature is visible.

Lemma 14.1 (Invertible preservation)**.** *If $Z = \phi(X)$ is invertible and sufficiently regular, then*

$$\kappa_T^Z(\phi(x)) = \kappa_T^X(x) \tag{94}$$

for every design contrast T.

Proof. Each transformed regime log density equals $h_s^X(x)$ plus the same log-Jacobian term. Every mixed design difference annihilates the common term. □

Lemma 14.2 (Design-sufficiency preservation)**.** *Let O be raw observation, $Z = \phi(O)$, and suppose*

$$S \perp O \mid Z. \tag{95}$$

Then, under the same state-independent pooling law, every posterior design odds ratio and hence every curvature field is preserved:

$$\kappa_T^O(o) = \kappa_T^Z(\phi(o)). \tag{96}$$

Proof. Design sufficiency gives $P(S = s \mid O = o) = P(S = s \mid Z = \phi(o))$. Curvature is a mixed difference of posterior log odds after subtracting known pooling odds, so the two representations agree. □

These lemmas turn the anti-gaming warning into a theorem. A representation may be audited only after it is shown to be invertible enough for the task or sufficient for the design label. The recommended sequence is:

1. learn an invertible or design-sufficient representation without rewarding flatness;
2. freeze it;
3. test whether raw observations retain regime information conditional on the representation;
4. test flatness on held-out data;
5. among representations meeting the same sufficiency threshold, prefer sparse local ratio supports.

This connects to mechanism-sparsity and intervention-based causal representation learning (Lachapelle et al., 2022; Ahuja et al., 2023; Kügelgen et al., 2023; Varici et al., 2025). In continuous coordinates,

$$\nabla_x \ell_e(x) = \nabla_x \log p_e(x) - \nabla_x \log p_0(x), \tag{97}$$

and the score difference is supported on the same intervened family as the potential. Score-based methods therefore study the differential of the object used here; the finite potential adds normalization, compositional products, and lattice curvature.

15 Simulation suite

The repository contains deterministic generators and exact-population checks for four scenarios.

15.1 Complete versus marginalized running example

With $a = 0.6$, $b = 0.5$, and $\sigma = 0.8$, the outcome synergy is $ab = 0.3$, complete-state curvature is zero, and marginal curvature ranges from

$$\log(1 - ab) = -0.356675 \tag{98}$$

to

$$\log(1 + ab) = 0.262364. \tag{99}$$

The scalar moment battery passes exactly.

15.2 Parity multi-pass failure

The balanced XOR mechanism in section 7.5 produces support $\{P, T\}$ and two invariant deletions. The intended output is not a forced orientation but the explicit status `AMBIGUOUS_MULTIPLE_PASSES`.

15.3 Latent conservation

Let independent $X, U \in \{-1, +1\}$, and set

$$L_A = 1 + aX + aU, \qquad L_B = 1 - aX + aU, \qquad 0 < a < 1/2. \tag{100}$$

Then

$$\mathbb{E}[L_A] = \mathbb{E}[L_B] = \mathbb{E}[L_A L_B] = 1. \tag{101}$$

After observing only X,

$$r_A = 1 + aX, \qquad r_B = 1 - aX, \qquad r_{AB} = 1, \tag{102}$$

and

$$\kappa = -\log(1 - a^2). \tag{103}$$

For $a = 0.3$, $\kappa = 0.094311$, $\mathrm{Cov}(r_A, r_B) = -0.09$, and $\mathbb{E}[\mathrm{Cov}(L_A, L_B \mid X)] = 0.09$.

These L_A, L_B are coordinated multi-source tilts. They satisfy complete-state multiplicative flatness, which is all that the marginalization identity requires, but they are not single-mechanism

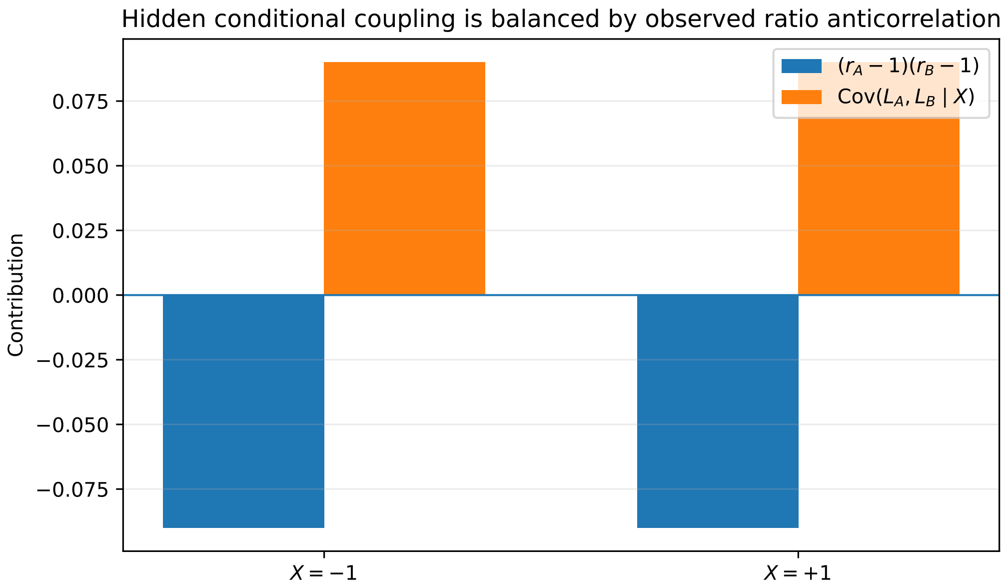

Figure 2: The latent conservation simulation. Pointwise conditional covariance of the complete factors is balanced by anticorrelation of the observed primitive ratios.

ratios on a simple DAG over independent (X, U). An on-model parent-centered construction can replace them when locality itself is the object of simulation.

15.4 Implementation inconsistency

Let the primitive ratios be $1 + aX_1$ and $1 + bX_2$, but let the combined intervention acquire an additional normalized implementation factor

$$\frac{1 + \gamma X_1 X_2}{1 + ab\gamma}. \tag{104}$$

Then

$$\kappa(X_1, X_2) = \log(1 + \gamma X_1 X_2) - \log(1 + ab\gamma). \tag{105}$$

With $a = 0.45$, $b = 0.35$, and $\gamma = 0.4$, curvature is -0.571921 when $X_1X_2 = -1$ and 0.275377 when $X_1X_2 = 1$. Setting $\gamma = 0$ is an exact negative control.

16 Empirical programs

16.1 Engineered feature flags

An engineered system offers the cleanest first validation because assignment, implementation, and telemetry can be controlled. Select two or more flags believed to alter distinct modules. Randomize them independently at the request, session, host, or deployment-unit level, and collect a rich state including local module outputs, queues, resource pressure, latency distributions, error traces, and

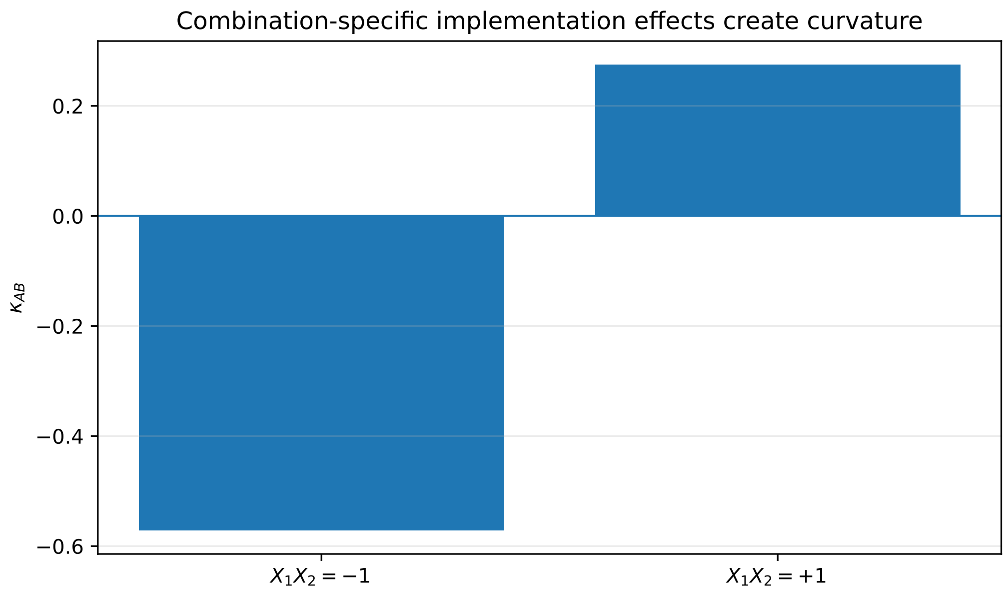

Figure 3: Combination-specific implementation effects create curvature even when the nominal primitive perturbations are conceptually distinct.

intervention-delivery metadata.

The protocol is:

1. verify product assignment and no state-dependent inclusion;
2. define negative-control flag pairs with separate implementation paths;
3. localize each primitive regime node;
4. run simultaneous deletion-equivalence orientation;
5. run GCM/wGCM projections and a joint multinomial interaction diagnostic;
6. hold out selected combinations and predict their full state distributions from primitive ratios;
7. add candidate telemetry when curvature is found and apply the nested state-expansion test;
8. cluster uncertainty at the true randomization unit.

This experiment can distinguish nonlinear product behavior from shared-resource coupling and from missing telemetry. Resolution-V fractions become useful once the number of flags exceeds what a full factorial can support.

16.2 Combinatorial Perturb-seq

Combinatorial Perturb-seq provides single and paired genetic perturbations with rich single-cell phenotypes (Norman et al., 2019). For each pair (A, B), fit the four-corner model

$$\eta_{ab}(x) = af_A(x) + bf_B(x) + abf_{AB}(x). \tag{106}$$

Estimate curvature at the biological-replicate level, not by treating cells as independent randomized units. Separate intervention-delivery curvature using alternative guides, guide-position swaps, expression or knockdown measurements, and non-targeting controls.

The central analysis is two-dimensional. Compare conventional genetic interaction summaries with full-distribution curvature. The decisive cases are perturbation pairs with strong outcome epistasis but low curvature, and pairs with weak selected-outcome interaction but high distributional curvature. Same-complex or same-pathway pairs may be enriched for curvature, but that is an empirical hypothesis, not a theorem.

Fit control and single perturbations, self-normalize the product ratio, and predict held-out paired distributions. Compositional perturbation predictors such as CPA impose additive latent structure (Lotfollahi et al., 2023); mechanism interferometry specifies where exact additivity belongs—the log density of a design-sufficient state—and supplies a test for when the prior fails.

17 Failure taxonomy and limitations

A nonzero curvature field proves that the proposed regime family is not closed under the asserted composition law. It does not uniquely identify the cause. Important possibilities are:

1. genuine mechanism coupling;
2. two perturbations sharing a target;
3. omitted common state;
4. state-dependent selection;
5. combination-specific implementation or dosage;
6. regime-dependent measurement;
7. poor overlap or estimation error;
8. incorrect correspondence between nominal and actual interventions.

Likewise, zero curvature is relative. It does not prove that no latent variables exist; it says that the chosen perturbations do not expose a failure of closure at the observed resolution.

The strongest graph-identification conclusions require ratio faithfulness, deletion faithfulness, adequate target coverage, and acyclicity. Conditional-independence testing remains difficult in high dimension (Shah and Peters, 2020). Hard interventions may fall outside the common-support calculus. In time series, whole-trajectory ratios can have variance that grows exponentially with horizon; estimation should occur at the transition level, pooling local conditional ratios with trajectory-level uncertainty. Static DAGs do not directly cover feedback at equilibrium, where Modular Response Analysis provides an important complementary tradition (Bruggeman et al., 2002; Kholodenko et al., 2002).

18 Relation to prior work and novelty

The locality result is the Markov blanket of a context node in an augmented causal graph, squarely connected to causal discovery from mechanism changes and Joint Causal Inference. The density ratio is classical change of measure and has recently been promoted as a pointwise causal object

(Mahadevan, 2026a). The exact pairwise product identity and four-corner null have direct precedent in Xu et al. (2024), including an empirical study using microscopy readouts from 50 gene-pair knockouts. That work strengthens the case that distribution-level four-corner tests are empirically viable on unstructured biological data.

The certificate also admits an exact sparsity reading. Under modularity, the master log-linear model (1) confines all 2^K regime log laws, pointwise in x, to the $(K+1)$-dimensional main-effects subspace spanned by the baseline log density and one potential per primitive perturbation; without modularity the family occupies up to 2^K unconstrained functions. Every nonvanishing curvature field is an additional function that the regime family forces the description to carry, and the Möbius expansion (13) enumerates these correction terms order by order. Modularity is therefore a maximal-parsimony statement about the regime family, and curvature terms are its epicycles. This is a bridge of principle to the algorithmic-independence-of-mechanisms and minimum-description-length traditions in causal inference (Janzing and Schölkopf, 2010; Grünwald, 2007), not a codelength theorem: the object constrained here is the exact dimension of a function-valued design subspace, derived from an explicit interventional algebra rather than postulated about observational complexity. The practical corollary is a tie-breaking principle used throughout the protocol: among localized supports and representations that meet the same held-out sufficiency threshold, prefer the most parsimonious one, because by locality the true support $\{t\} \cup \mathrm{pa}(t)$ is, under ratio and deletion faithfulness and adequate data, the smallest support consistent with the regime information.

The differential-geometric program of Mahadevan (2026b) is conceptually adjacent: failures of intervention-induced fields to close can witness unmodeled structure. Its Lie brackets measure sequential infinitesimal noncommutativity, while the present construction is a finite symmetric mixed difference on a Boolean design and requires no smooth intervention path.

No individual probability identity in this paper should be claimed as novel. The narrow contribution is the assembled object:

$$\boxed{\text{local family support} + \text{targetwise normalization} + \text{gauge-invariant square flatness}} \tag{107}$$

$$\boxed{\iff \text{existence of a context-invariant modular soft-intervention model,}} \tag{108}$$

together with ratio-free orientation, support-error diagnostics, exact latent conservation, missing-information interpretation, state-expansion conservation, and the dual inference architecture. This is a literature-based novelty assessment, not proof of priority.

19 Conclusion

Mechanism interferometry treats a factorial family of empirical regimes as one constrained mathematical object. The central logic can be summarized in four lines:

$$\text{locality} \implies \text{which causal family changed,} \tag{109}$$

$$\text{deletion invariance} \implies \text{which family member is the target,} \tag{110}$$

$$\text{square curvature} \implies \text{where autonomous composition fails,} \tag{111}$$

$$\text{conditional score covariance} \implies \text{how much joint mechanism information was lost.} \tag{112}$$

The method separates system nonlinearity, genuine mechanism coupling, and representational incompleteness. Evidence for causality no longer rests on one significant coefficient. It rests on simultaneous satisfaction of locality, normalization, square closure, moment restrictions, held-out compositional prediction, overlap requirements, and implementation negative controls. The resulting overdetermination is the source of evidential strength.

A Additional proofs

A.1 Square flatness implies all-order flatness

Fix x and suppress it. Let $h : \{0,1\}^K \to \mathbb{R}$. For coordinate j, define $\delta_j(s) = h(s + e_j) - h(s)$ wherever $s_j = 0$. If every square contrast vanishes, then for every $k \neq j$,

$$\delta_j(s + e_k) - \delta_j(s) = \Delta_k \Delta_j h(s) = 0. \tag{113}$$

Every two vertices of the subcube $s_j = 0$ are connected by flips of coordinates other than j. Hence $\delta_j(s) = \delta_j(0)$. Summing along a monotone path from 0 to s gives

$$h(s) = h(0) + \sum_j s_j \delta_j(0). \tag{114}$$

Every mixed difference of order at least two annihilates this affine form.

A.2 Peeling reconstruction from unlabeled families

Let $\mathcal{F} = \{F_i : i \in [d]\}$ with $F_i = \{i\} \cup \mathrm{pa}(i)$. Maintain identified set U. If $U \neq [d]$, choose the earliest node $i \notin U$ in a topological ordering restricted to unidentified nodes. Every parent of i lies in U, so

$$F_i \setminus U = \{i\}. \tag{115}$$

Conversely, if an unassigned family reduces to $\{i\}$, its target must be i, because every family contains its own target and no other unidentified element. Assign the family, recover $\mathrm{pa}(i) = F_i \setminus \{i\}$, add i to U, and iterate. This proves uniqueness.

A.3 Preservation under design sufficiency

For any two corners s, t, Bayes' rule under pooling law ρ gives

$$\log \frac{p_s(o)}{p_t(o)} = \log \frac{P_\rho(S = s \mid O = o)}{P_\rho(S = t \mid O = o)} - \log \frac{\rho_s}{\rho_t}. \tag{116}$$

If $S \perp O \mid Z$, the posterior odds depend on o only through $z = \phi(o)$. Every linear contrast of regime log ratios, including all Möbius curvatures, is therefore preserved.

A.4 Nested state expansion

By the conditional law of total covariance,

$$\begin{aligned} \mathrm{Cov}(L_A, L_B \mid X) = \mathbb{E}[\mathrm{Cov}(L_A, L_B \mid X, W) \mid X] \quad (117)\\ + \mathrm{Cov}(\mathbb{E}[L_A \mid X, W], \mathbb{E}[L_B \mid X, W] \mid X). \quad (118) \end{aligned}$$

Substitute

$$\mathrm{Cov}(L_A, L_B \mid X) = r_A^X r_B^X (e^{\kappa_X} - 1) \tag{119}$$

and the analogous identity at (X, W).

B Partial-design estimability

Let $D = \{s^{(1)}, \ldots, s^{(n_D)}\}$. Define the main-effects matrix

$$M_D = \begin{bmatrix} 1 & s_1^{(1)} & \cdots & s_K^{(1)} \\ \vdots & \vdots & & \vdots \\ 1 & s_1^{(n_D)} & \cdots & s_K^{(n_D)} \end{bmatrix}. \tag{120}$$

At each x, the flat model is $h_D(x) = M_D \beta(x)$. A contrast $c^\top h_D(x)$ is a testable implication of flatness exactly when $c \in \ker(M_D^\top)$. The dimension of the lack-of-fit space is

$$n_D - \mathrm{rank}(M_D). \tag{121}$$

An implementation should compute a numerically stable basis of this null space, map familiar face contrasts into it, and report which directions are untested or aliased. For binary fractions, standard defining relations and generalized word-length patterns can be carried over pointwise.

The six-corner design $\{0, 1\}^3 \setminus \{000, 111\}$ has $n_D = 6$ and rank four. It has two independent lack-of-fit contrasts despite containing no complete square face. This demonstrates why "test all observed squares" is complete only under a design-topology condition.

C Simulation derivations

C.1 Gaussian ratio identity

For $\phi_\sigma(z) = (2\pi\sigma^2)^{-1/2} \exp(-z^2/(2\sigma^2))$,

$$\frac{\phi_\sigma(y-1)}{\phi_\sigma(y+1)} = \exp\left(\frac{2y}{\sigma^2}\right). \tag{122}$$

Writing $R = e^{2y/\sigma^2}$,

$$\frac{\phi_\sigma(y-1) - \phi_\sigma(y+1)}{\phi_\sigma(y-1) + \phi_\sigma(y+1)} = \frac{R-1}{R+1} = \tanh(y/\sigma^2). \tag{123}$$

C.2 Parity arithmetic

Under either mechanism, P is uniform. For each value of T, summing over P gives probability 1/2, regardless of ε. Hence both one-dimensional marginals are invariant, even though the conditional relation between P and T reverses.

C.3 Implementation factor normalization

Under primitive tilts, $\mathbb{E}^{\text{flat}}_{AB}[X_1X_2] = ab$. Therefore

$$\mathbb{E}^{\text{flat}}_{AB}[1 + \gamma X_1X_2] = 1 + ab\gamma, \tag{124}$$

which proves that the factor used in the implementation-inconsistency simulation is normalized.

D Longitudinal extension

For a transition model,

$$p_e(x_{0:T}) = p_e(x_0) \prod_{t=1}^{T} p_e(x_t \mid x_{<t}). \tag{125}$$

If only one transition mechanism changes, the log trajectory ratio is a sum of local transition log ratios. Direct products can have exponentially growing second moments. The implementation should therefore estimate transition-level ratios, pool across comparable transitions, and use trajectory- or subject-level resampling. Feedback is represented by the unrolled temporal graph rather than by pretending that a static DAG contains no cycles.

E Software audit contract

Every empirical run should emit a machine-readable audit bundle containing:

1. input fingerprints and schema version;
2. regime definitions, pooling proportions, and product-sampling check;

3. cross-fitting fold assignments and deterministic random seeds;
4. selection and overlap diagnostics;
5. localization stability paths;
6. simultaneous deletion-equivalence intervals and pass status;
7. exploratory proposal artifacts, their candidate-library fingerprints, tie rules, and recommendation-or-abstention status;
8. moment battery, GCM/wGCM projections with their serialized product-design evidence, and joint-model diagnostics;
9. estimator-family agreement results across the declared lens battery;
10. raw and self-normalized compositional predictions;
11. state-expansion comparisons;
12. negative-control results;
13. all warnings, abstentions, and strict-mode failures.

The companion repository defines JSON schemas for these artifacts and a Rust workspace architecture for producing them.

References

Ahuja, Kartik, Divyat Mahajan, Yixin Wang, and Yoshua Bengio (2023). "Interventional Causal Representation Learning". In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. Proceedings of Machine Learning Research. PMLR, pp. 372–407. URL: https://proceedings.mlr.press/v202/ahuja23a.html.

Box, George E. P., J. Stuart Hunter, and William G. Hunter (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*. 2nd ed. Wiley.

Brillinger, David R. (1969). "The Calculation of Cumulants via Conditioning". In: *Annals of the Institute of Statistical Mathematics* 21, pp. 215–218. DOI: 10.1007/BF02532246.

Brouillard, Philippe, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin (2020). "Differentiable Causal Discovery from Interventional Data". In: *Advances in Neural Information Processing Systems*. Vol. 33, pp. 21865–21877.

Bruggeman, Frank J., Hans V. Westerhoff, Jan B. Hoek, and Boris N. Kholodenko (2002). "Modular Response Analysis of Cellular Regulatory Networks". In: *Journal of Theoretical Biology* 218.4, pp. 507–520. DOI: 10.1006/jtbi.2002.3096.

Chen, Hua Yun (2007). "A Semiparametric Odds Ratio Model for Measuring Association". In: *Biometrics* 63.2, pp. 413–421. DOI: 10.1111/j.1541-0420.2006.00701.x.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018). "Double/Debiased Machine Learning for Treatment and Structural Parameters". In: *The Econometrics Journal* 21.1, pp. C1–C68. DOI: 10.1111/ectj.12097.

Eaton, Daniel and Kevin Murphy (2007). "Exact Bayesian Structure Learning from Uncertain Interventions". In: *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*. Vol. 2. Proceedings of Machine Learning Research. PMLR, pp. 107–114. URL: https://proceedings.mlr.press/v2/eaton07a.html.

Greenland, Sander, James M. Robins, and Judea Pearl (1999). "Confounding and Collapsibility in Causal Inference". In: *Statistical Science* 14.1, pp. 29–46. DOI: 10.1214/ss/1009211805.

Gretton, Arthur, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola (2012). "A Kernel Two-Sample Test". In: *Journal of Machine Learning Research* 13, pp. 723–773. URL: https://jmlr.org/papers/v13/gretton12a.html.

Grünwald, Peter D. (2007). *The Minimum Description Length Principle*. MIT Press.

Janzing, Dominik and Bernhard Schölkopf (2010). "Causal Inference Using the Algorithmic Markov Condition". In: *IEEE Transactions on Information Theory* 56.10, pp. 5168–5194. DOI: 10.1109/TIT.2010.2060095.

Kholodenko, Boris N., Anatoly Kiyatkin, Frank J. Bruggeman, Eduardo Sontag, Hans V. Westerhoff, and Jan B. Hoek (2002). "Untangling the Wires: A Strategy to Trace Functional Interactions in Signaling and Gene Networks". In: *Proceedings of the National Academy of Sciences* 99.20, pp. 12841–12846. DOI: 10.1073/pnas.192442699.

Kügelgen, Julius von, Michel Besserve, Wenbo Liang, Luigi Gresele, Armin Kekić, Elias Bareinboim, David M. Blei, and Bernhard Schölkopf (2023). "Nonparametric Identifiability of Causal Representations from Unknown Interventions". In: *Advances in Neural Information Processing Systems*. Vol. 36, pp. 48603–48638.

Lachapelle, Sebastien, Pau Rodriguez, Yash Sharma, Katie E. Everett, Rémi Le Priol, Alexandre Lacoste, and Simon Lacoste-Julien (2022). "Disentanglement via Mechanism Sparsity Regularization: A New Principle for Nonlinear ICA". In: *Proceedings of the First Conference on Causal Learning and Reasoning*. Vol. 177. Proceedings of Machine Learning Research. PMLR, pp. 428–484. URL: https://proceedings.mlr.press/v177/lachapelle22a.html.

Lotfollahi, Mohammad, Mohsen Naghipourfar, Fabian J. Theis, and F. Alexander Wolf (2023). "Predicting Cellular Responses to Complex Perturbations in High-Throughput Screens". In: *Molecular Systems Biology* 19.6, e11517. DOI: 10.15252/msb.202211517.

Louis, Thomas A. (1982). "Finding the Observed Information Matrix when Using the EM Algorithm". In: *Journal of the Royal Statistical Society: Series B* 44.2, pp. 226–233. DOI: 10.1111/j.2517-6161.1982.tb01203.x.

Mahadevan, Sridhar (2026a). "Causal Density Functions". In: *arXiv preprint arXiv:2606.00754*. arXiv: 2606.00754.

— (2026b). "Infinitesimal Causality". In: *arXiv preprint arXiv:2606.24621*. arXiv: 2606.24621.

Meinshausen, Nicolai and Peter Bühlmann (2010). "Stability Selection". In: *Journal of the Royal Statistical Society: Series B* 72.4, pp. 417–473. DOI: 10.1111/j.1467-9868.2010.00740.x.

Mooij, Joris M., Sara Magliacane, and Tom Claassen (2020). "Joint Causal Inference from Multiple Contexts". In: *Journal of Machine Learning Research* 21.99, pp. 1–108. URL: https://jmlr.org/papers/v21/17-123.html.

Norman, Thomas M., Max A. Horlbeck, Joseph M. Replogle, Alex Y. Ge, Albert Xu, Marco Jost, Luke A. Gilbert, and Jonathan S. Weissman (2019). "Exploring Genetic Interaction Manifolds Constructed from Rich Single-Cell Phenotypes". In: *Science* 365.6455, pp. 786–793. DOI: 10.1126/science.aax4438.

Scheidegger, Cyrill, Julia Hörrmann, and Peter Bühlmann (2022). "The Weighted Generalised Covariance Measure". In: *Journal of Machine Learning Research* 23.273, pp. 1–68. URL: https://jmlr.org/papers/v23/21-1328.html.

Shah, Rajen D. and Jonas Peters (2020). "The Hardness of Conditional Independence Testing and the Generalised Covariance Measure". In: *The Annals of Statistics* 48.3, pp. 1514–1538. DOI: 10.1214/19-AOS1857.

Shirts, Michael R. and John D. Chodera (2008). "Statistically Optimal Analysis of Samples from Multiple Equilibrium States". In: *The Journal of Chemical Physics* 129.12, p. 124105. DOI: 10.1063/1.2978177.

Squires, Chandler, Yuhao Wang, and Caroline Uhler (2020). "Permutation-Based Causal Structure Learning with Unknown Intervention Targets". In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*. Vol. 124. Proceedings of Machine Learning Research. PMLR, pp. 1039–1048. URL: https://proceedings.mlr.press/v124/squires20a.html.

Székely, Gábor J. and Maria L. Rizzo (2013). "Energy Statistics: A Class of Statistics Based on Distances". In: *Journal of Statistical Planning and Inference* 143.8, pp. 1249–1272. DOI: 10.1016/j.jspi.2013.03.018.

Tchetgen Tchetgen, Eric J., James M. Robins, and Andrea Rotnitzky (2010). "On Doubly Robust Estimation in a Semiparametric Odds Ratio Model". In: *Biometrika* 97.1, pp. 171–180. DOI: 10.1093/biomet/asp062.

Tian, Jin and Judea Pearl (2001). "Causal Discovery from Changes". In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, pp. 512–521.

Tsamardinos, Ioannis, Constantin F. Aliferis, and Alexander Statnikov (2003). "Algorithms for Large Scale Markov Blanket Discovery". In: *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference*. AAAI Press, pp. 376–381.

VanderWeele, Tyler J. and Mirjam J. Knol (2014). "A Tutorial on Interaction". In: *Epidemiologic Methods* 3.1, pp. 33–72. DOI: 10.1515/em-2013-0005.

Varici, Burak, Emre Acartürk, Karthikeyan Shanmugam, and Ali Tajer (2025). "Score-Based Causal Representation Learning with Interventions". In: *Journal of Machine Learning Research*. See also arXiv:2301.08230.

Wu, C. F. Jeff and Michael S. Hamada (2009). *Experiments: Planning, Analysis, and Optimization*. 2nd ed. Wiley.

Xu, Zuheng, Moksh Jain, Ali Denton, Shawn Whitfield, Aniket Didolkar, Berton Earnshaw, and Jason Hartford (2024). "Automated Discovery of Pairwise Interactions from Unstructured Data". In: *arXiv preprint arXiv:2409.07594*. arXiv: 2409.07594.