PERCEPTIVE CLOSURE: IDENTIFYING AUTHORIZATION-RESOURCE COUNTERFACTUALS

Anonymous authors
Paper under double-blind review

ABSTRACT

Resource-specific claims about governed authorization—for example that representation is unnecessary, substitutable, or structurally required—are themselves counterfactual claims. We ask when such claims are identified. For a sealed authorization target, PERCEPTIVE CLOSURE (PC) represents evidence admission (E), certificate representation (R), and epistemic authority (A) as extensional contract coordinates and evaluates closure under every subset freeze. A complete contract yields a resource-freezing signature K_Π; source/specification evidence Ω yields the compatible identified set $\mathcal{K}_{\Pi,\Omega}$. We prove three stages. First, unrestricted controller behavior need not identify resource-freezing dependence: behaviorally equivalent systems can have arbitrary finite differences, or finite-versus-structural differences, under an R freeze. Second, one fixed implementation description can admit source-consistent E- and R-typed completions with different freeze values. Third, conditional on explicitly declared semantic anchors and the source-exposed intervention boundary, touch is mechanically determined; interface-equivariant resource-support bisimulation is a sufficient certificate for invariance across interface-preserving refactorings. We then show that this identification boundary is operational rather than merely constructed. Apache Polaris is naturally E/R boundary-underidentified, whereas a pre-existing XACML 3.0/AuthzForce 21.2.0 environment point-identifies an evidence-touching repair with

$$T(q) = \{E\}, \qquad K_\Pi = [1, \infty, 1, 1, \infty, \infty, 1, \infty], \qquad |\mathcal{K}_{\Pi,\Omega}| = 1.$$

The XACML policy and native engine are unchanged and no manual $E/R/A$ action labels are used. PC-TAU supplies controlled nondegenerate freeze geometry, and live tool-use episodes establish finite-fallback realizability. PC does not claim prevalence, universal safety, causal decomposition inside mixed actions, or a universal $E/R/A$ ontology. Its claim is narrower: planning evaluates closure under a specified resource counterfactual; PC asks whether that resource counterfactual is warranted by the declared semantics in the first place.

1 INTRODUCTION

Resource-specific authorization diagnoses are inferential claims. Saying that closure "requires representation" asserts what would happen if R-touching interventions were unavailable. Saying that evidence is substitutable or authority is the bottleneck makes the same kind of intervention-sensitive assertion. Such claims can distinguish zero dependence, finite fallback, structural loss of closure, and cross-resource complementarity, but only relative to an identified intervention interface.

Neither complete unrestricted behavior nor complete implementation dataflow necessarily determines which governance resource closure depends on.

PC therefore separates two questions that are easy to conflate:

identify the authorization-resource counterfactual
$\Downarrow$
compute closure under that counterfactual

The second problem is meaningful only after the first is settled. The distinction matters whenever the same raw fact can occupy different governance roles. A field can be present in runtime memory but

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.

absent from normalized authorization-admitted history; admitted but aliased by the certificate-facing representation; represented but unusable because the authority state does not license the relevant interface; or modified by an atomic action that changes several coordinates at once.

Authorization repair and information acquisition are not new. UCON permits mutable authorization-relevant state; automated trust negotiation and state-modifying authorization reason over disclosures and authorization-changing action sequences; dynamic-policy logics reason over stateful policy behavior and equivalence; Margrave verifies XACML policies and computes semantic differences between supplied policy versions; decision-region determination acquires information until all remaining hypotheses imply one decision; and causal abstraction/partial-identification theory already separates observational information from intervention-sensitive or set-identified structure Park and Sandhu (2004); Winsborough and Li (2006); Becker and Nanz (2007); Dougherty et al. (2006); Fisler et al. (2005); Javdani et al. (2014); Beckers and Halpern (2019); Kline and Tamer (2023); Li et al. (2025). PC does not claim novelty for any of those ingredients.

Its object is narrower: *authorization-resource intervention identification*. For a complete declared contract, K_Π describes closure under resource freezes. From source evidence Ω, $\mathcal{K}_{\Pi,\Omega}$ retains every compatible signature instead of silently selecting one semantic completion. This turns a familiar what-if question on its head. Change-impact reasoning asks what changes *given* a policy or intervention change. PC asks whether the available source evidence identifies which resource intervention is licensed as the counterfactual in the first place.

The paper's strongest consequence is therefore a boundary result, not a new planning algorithm:

$$\begin{aligned} \text{unrestricted behavior} &\not\Rightarrow \text{resource counterfactual,} \\ \text{implementation/dataflow} &\not\Rightarrow \text{resource counterfactual,} \\ \text{source-identified semantic anchors} &\Rightarrow \text{resource counterfactual can be point-identified.} \end{aligned}$$

Our central empirical observation is that independently engineered systems occupy both sides of this boundary. Apache Polaris exposes a real historical authorization repair whose raw source admits both E- and R-consistent semantic completions. In contrast, a pre-existing XACML/AuthzForce fixture supplies sufficient external semantics to fix the measured interface and yields a singleton PC signature. The positive case is deliberately narrow: the external standard/system anchors the semantic interface and native request/response transaction, while the finite two-world measurement wrapper, cost normalization, freeze order, and recording apparatus are experiment-authored.

Contributions.

1. **Identification object.** We define extensional $E/R/A$ touch from a complete authorization contract and the source-relative identified set $\mathcal{K}_{\Pi,\Omega}$. We prove that unrestricted behavior does not determine even the sign or finiteness of R-freezing dependence.
2. **Source insufficiency and conditional recovery.** One fixed implementation/dataflow description can admit E- and R-typed completions with arbitrarily different freeze values. When H, P_R, Λ and the source intervention boundary are fixed extensionally, touch becomes mechanical, and resource-support bisimulation is a sufficient interface-preserving certificate for equality of every freeze value.
3. **Natural identification boundary.** Polaris naturally remains E/R boundary-underidentified, while XACML/AuthzForce naturally point-identifies a pure E-touching Stage-III witness with a singleton identified set.
4. **Geometry after identification.** We retain exact rectangular partial-identification bounds, zero/finite/structural criteria, strict E/R complementarity, controlled finite substitution in PC-TAU, and learned-agent fallback realizability. These characterize consequences once the counterfactual is identified rather than replacing the identification question.

2 THE AUTHORIZATION-RESOURCE OBJECT

Fix finite latent worlds X and a sealed target $A_\Pi : X \to D$. A finite complete PC contract is

$$C = (X, U_H, \sim, S, s_0, H, P_R, \Lambda, \omega, Q, \mathrm{Succ}^+, Y, c, A_\Pi, \mathrm{Atom}),$$

where $\sim \subseteq X \times U_H$ is world–history compatibility, $s_0 \in S$, $H : S \to U_H$, $P_R : S \to \mathrm{Eq}(U_H)$, $\Lambda : S \to \mathfrak{A}$, $\omega : S \to O$, and $A_\Pi : X \to D$; $\mathfrak{A}, O, D$ are the authority, observation, and target

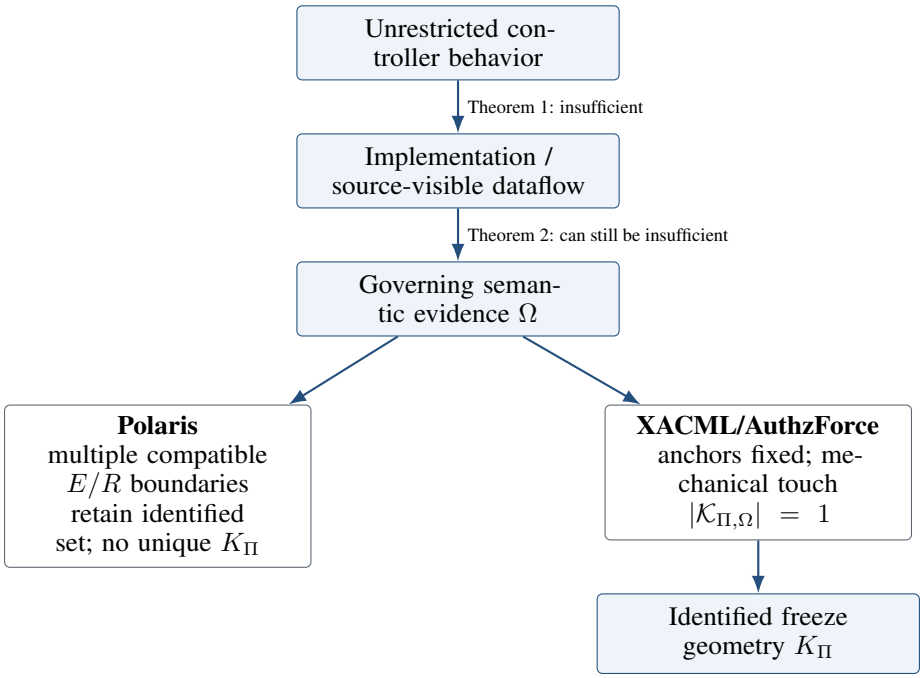

Figure 1: Identification boundary. PC does not assume that source code determines the resource counterfactual. The governing question is whether all source-compatible semantic completions agree on the resulting freeze signature. Natural external systems can fall on either side.

codomains. Each $Q(s)$ is a finite set of source-atomic interventions, $\mathrm{Succ}^+(s,q) \subseteq S$ is nonempty for enabled q, Y gives controller-visible outcome classes, $c(q) \geq 0$, and Atom fixes the source action/checkpoint boundary. Unqualified system-level κ_Π, K_Π, and $T(q)$ below are evaluated at s_0 when the action identity is unambiguous. Well-formedness requires $P_R(s)$ to be an equivalence relation, reachable $C_R(s)$ (defined next) to be nonempty, and atomicity/successors to agree with that interface. With current certificate cell $[H(s)]_{P_R(s)}$,

$$C_R(s) = \{x : \exists h \in [H(s)]_{P_R(s)} \text{ with } x \sim h\}.$$

Authority coherence is extensional: if $Q_{\mathrm{src}}(s)$ is the source-available candidate set, $Q(s)$ contains exactly candidates whose ordinary precondition holds and whose interface is admitted by $\Lambda(s)$; holding non-authority state/interface fixed, a change in authority-admissibility must change Λ.

Definition 1 (Target-relative closure)**.** *A checkpoint s is closed iff $|\{A_\Pi(x) : x \in C_R(s)\}| = 1$. It is open otherwise. A closed checkpoint licenses only that target-relative authorization class; an open checkpoint may be repaired or externally escalated.*

For each action $q \in Q(s)$ and each positive-support successor $s' \in \mathrm{Succ}^+(s,q)$, define canonical resource touch

$$T_s(q) = \bigcup_{s' \in \mathrm{Succ}^+(s,q)} \{E : H(s') \neq H(s),\ R : P_R(s') \neq P_R(s),\ A : \Lambda(s') \neq \Lambda(s)\}. \quad (1)$$

Equality of P_R and Λ is extensional: renaming a representation value does not touch R, and satisfying an unchanged authority rule does not touch A. In particular, the authority-coherence condition prevents an authority-only availability change from hiding in Q while Λ remains extensionally equal. H is normalized authorization-admitted history, not raw controller scratch state. Mixed source-atomic actions keep the union of effects. A freeze never fabricates a partial subaction that the source does not expose.

Action availability, not internal causal decomposition. Equation (1) is deliberately transition-semantic. Because Succ^+ contains only positive-support semantic successors, changes that never occur on a source-valid successor do not generate touch. But if a reachable source-atomic action

genuinely changes P_R or Λ, that coordinate is touched even when the changed component is not minimally necessary for the current target. A resource freeze removes resource-touching *actions*; PC does not invent a hypothetical subaction that preserves only the target-relevant internal effect. This is why Δ_R below is an R-freezing penalty rather than a causal attribution score inside a mixed action.

Interface-relativity witness. Consider one underlying transaction with an evidence-admission phase e that already makes the represented fiber target-homogeneous, followed by an R-side effect r that refreshes a representation relation without changing the target decision. If the governing interface exposes the transaction only as one atomic action q_{ER}, then $T(q_{ER}) = \{E, R\}$ and an R freeze removes the only closer. If the governing interface instead exposes e independently, then the E-only closer can survive an R freeze. The end-to-end authorization decision can be identical while K_Π differs. PC therefore does not claim a universally canonical decomposition: it uses the source-exposed intervention interface fixed by Ω; if Ω does not determine that interface, the resulting signatures remain set-identified rather than being resolved by analyst preference.

The $E/R/A$ decomposition is therefore an extensional interface factorization, not a manual action label or universal ontology. Within the declared contract class it is componentwise irredundant: H, P_R, Λ can vary separately, Proposition 2 gives finite erasure witnesses, and Proposition 6 shows strict E/R complementarity. Crucially, K_Π is defined relative to the declared source action/checkpoint interface. Split/merge, batching, transactional decomposition, or checkpoint moves that change which semantic effects are independently source-exposed may change the estimand; Theorem 3 does not claim invariance to such interface-changing refactors.

A proper closing policy tree is a finite ω-measurable contingent tree whose every positive-support leaf is closed. Under worst-positive-support evaluation, randomized policies do not improve the value: replacing a supported distribution by a least-valued supported action cannot increase worst-branch cost. Let $\mathcal{R}_\Pi(s)$ denote proper closing policies and define

$$\kappa_\Pi(s) = \inf_{\rho \in \mathcal{R}_\Pi(s)} \max_{\tau \in \mathrm{Leaves}(\rho)} \sum_{q \in \tau} c(q), \tag{2}$$

with $\kappa_\Pi(s) = \infty$ if no proper closer exists. For a resource-freeze set $F \subseteq G := \{E, R, A\}$, the freeze $\kappa_\Pi^{\neg F}$ removes every source-exposed atomic action q with $T_s(q) \cap F \neq \varnothing$. The full signature is

$$K_\Pi(s) = \left(\kappa_\Pi^{\neg F}(s)\right)_{F \subseteq G}, \tag{3}$$

with coordinate order $(\varnothing, E, R, A, ER, EA, RA, ERA)$. For finite $\kappa_\Pi(s)$ define $\Delta_R := \kappa_\Pi^{\neg\{R\}} - \kappa_\Pi \in [0, \infty]$, with $\infty - a = \infty$; values 0, $(0, \infty)$, and ∞ denote zero, finite-positive, and structural dependence. If unrestricted closure is infinite, Δ_R is undefined and the completion is baseline-unclosable.

Proposition 1 (Control–certificate separation)**.** *There are finite PC instances for which no single untyped history quotient is simultaneously (i) complete enough to support a budget-one zero-error repair policy and (ii) faithful enough to serve as the terminal authorization certificate partition.*

Appendix A uses controller-distinguishable but certificate-aliased checkpoints, so controller choice cannot serve as terminal evidence.

3 Three-stage identification theory

3.1 Stage I: behavior is not enough

Let $\mathrm{Beh}(M, \rho)$ be the rooted positive-support tree induced by unrestricted policy ρ, labeled by controller observation, source action identity, controller-visible outcome, cost, terminal class, and stopping behavior.

Definition 2 (Unrestricted controller equivalence)**.** *Write $M \equiv_{\mathrm{ctrl}} M'$ when the systems have the same declared controller interface and, for every unrestricted ω-measurable policy, the induced labeled support trees are isomorphic in both directions.*

Theorem 1 (Behavioral non-identifiability of R-freezing dependence)**.** *For every finite $g > 0$ there exist finite complete PC systems M_g, M_g' with $M_g \equiv_{\mathrm{ctrl}} M_g'$ and unrestricted value 1 in both, but*

$$\kappa_{\Pi, M_g}^{\neg\{R\}} = 1 + g, \qquad \kappa_{\Pi, M_g'}^{\neg\{R\}} = 1.$$

Hence $\Delta_R = g$ versus 0. There is also a behaviorally equivalent pair with unrestricted value 1 and frozen-R value ∞ versus 1. Thus neither K_Π nor the sign/finiteness of R-freezing dependence is a functional of unrestricted controller behavior.

Diagnostic consequence. Any attribution determined only by unrestricted controller behavior must agree on the witness pair even though their correct R-freeze classifications differ. Appendix B gives the proof. The generic observation/intervention distinction is not claimed as new.

3.2 Stage II: incomplete source semantics are not enough

Definition 3 (Source evidence and semantic completions)**.** *Let Ω constrain source dataflow, target, source locators, action/checkpoint boundaries, and declared authorization semantics, with nonempty $\mathcal{C}(\Omega) = \{C : C \models \Omega\}$ and distinguished source entry ℓ_0. Each completion realizes source checkpoints by $\iota_C^S : \mathcal{L}_\Omega \to S_C$ and every action locator a fixed by Ω by $\iota_C^Q(\ell, a) \in Q_C(\iota_C^S(\ell))$. Then*

$$\mathcal{K}_{\Pi,\Omega} = \{K_\Pi^C(\iota_C^S(\ell_0)) : C \in \mathcal{C}(\Omega)\}, \qquad \mathcal{T}_{\Omega,\ell}(a) = \{T^C_{\iota_C^S(\ell)}(\iota_C^Q(\ell, a)) : C \in \mathcal{C}(\Omega)\}.$$

Definite/possible touch are the intersection/union of the latter. If Ω does not identify a common action/interface locator, that boundary remains underidentified rather than being compared through nonexistent common s, q indices.

Theorem 2 (Source-semantic boundary non-identifiability)**.** *For every finite $g > 0$ there exists a single finite implementation description Ω_g and two complete contracts $C_E, C_R \in \mathcal{C}(\Omega_g)$ that agree on the implementation graph, raw source datum, source action identities, controller observations/outcomes, costs, and target classes; they are unrestricted-controller equivalent with value 1, yet the same fast transition satisfies*

$$T^{C_E}(q) = \{E\}, \qquad T^{C_R}(q) = \{R\},$$

and

$$\kappa^{\neg\{R\}}_{\Pi,C_E} = 1, \qquad \kappa^{\neg\{R\}}_{\Pi,C_R} = 1 + g.$$

Deleting the common backup action yields a structural pair 1 versus ∞. Therefore source code/dataflow that omits the authorization-admission/certificate boundary need not identify the resource intervention or K_Π.

A source/specification record need not declare when a raw datum becomes authorization-admitted, what equality relation the certificate uses, or what future interfaces an authority state licenses. PC supplies no default anchor for an omitted boundary. The proof uses one raw context datum already present before an authorization-interface call: one completion treats it as already admitted and makes the call a certificate refinement; another treats admission as occurring at the call while keeping the certificate map fixed. The implementation is unchanged, but the freeze value changes by arbitrary finite g and, without the backup, from finite closure to structural dependence. When $\mathcal{K}_{\Pi,\Omega}$ is non-singleton, selecting one scalar resource signature adds semantics not fixed by Ω.

Corollary 1 (Specified-counterfactual evaluation does not imply counterfactual identification)**.** *There exists source evidence Ω for which the consequences of each supplied complete resource interpretation are exactly evaluable, while Ω itself does not identify which resource interpretation is warranted. Thus a sound procedure for "given intervention ι, what changes?" need not solve "does Ω identify which resource intervention ι is licensed?"*

Proof. Apply any exact evaluator separately to C_E and C_R in Theorem 2. It can correctly compute the two distinct supplied counterfactuals, but no function of Ω_g alone can select between them without adding an omitted semantic anchor. □

Corollary 1 is the formal separation from policy-difference and ordinary what-if analysis used in our novelty claim. Margrave, for example, semantically compares supplied XACML policy versions and reports inputs whose decisions differ Fisler et al. (2005). Dynamic-policy logics and state-modifying authorization can reason exactly about behavior, equivalence, and action sequences once their policy/state semantics are specified Dougherty et al. (2006); Becker and Nanz (2007). Those are complementary downstream capabilities. Theorem 2 shows why correct evaluation of a supplied counterfactual does not, by itself, identify the resource counterfactual from underspecified source evidence.

3.3 STAGE III: CONDITIONAL RECOVERY UNDER DECLARED ANCHORS

Stage III is conditional identification, not an extraction theorem from arbitrary source code. If Ω does not explicitly fix H, P_R, Λ, or the source-exposed successor/atomicity boundary, those quantities remain free in $\mathcal{C}(\Omega)$ and PC returns partial or boundary underidentification. Once the governing specification or audited contract fixes them extensionally, the complete contract can be reduced without retaining all implementation state.

PC requires source identification, not universal canonicality.

Define the resource-support kernel $\mathsf{K}(C)$ over semantic checkpoints: terminal target class, controller observation ω, source-exposed atomic action classes, canonical touch, nonnegative cost, controller-visible semantic outcome classes, and positive-support successors. Cycles and nontermination support are retained. For each observation o, an interface correspondence gives a bijection α_o between enabled source-atomic action classes and, for each (o, q), a bijection $\beta_{o,q}$ between positive-support outcome classes. A relation B is an interface-equivariant resource-support bisimulation when related checkpoints preserve terminal/open status, closed target class, observation, action cost and touch, and supported successors symmetrically. Syntactic renaming and hidden implementation plumbing may vary; split/merge, batching, transactional decomposition, or checkpoint refactors that alter the source-atomic action domain are outside the equivalence.

Theorem 3 (Conditional anchored identification and interface-preserving resource-kernel invariance)**.** *For finite PC contracts: (i) fixing H, P_R, Λ and the source-exposed successor/atomicity boundary uniquely determines every $E/R/A$ touch set by Equation* (1)*; (ii) if $\mathsf{K}(C)$ and $\mathsf{K}(C')$ are interface-equivariantly resource-support bisimilar at their initial checkpoints, then*

$$\kappa^{\neg F}_{\Pi,C} = \kappa^{\neg F}_{\Pi,C'} \qquad \text{for every } F \subseteq G,$$

so K_Π, Δ_R whenever the unrestricted value is finite, baseline-unclosability, and every zero/finite/infinite freeze classification agree; (iii) if every completion in $\mathcal{C}(\Omega)$ lies in one such kernel-equivalence class at the realized source entry $\iota^S_C(\ell_0)$, then $|\mathcal{K}_{\Pi,\Omega}| = 1$. Consequently, interface-preserving source refactorings cannot change the identified PC signature.

The theorem gives a sufficient structural certificate for equality of K_Π; it is not claimed necessary, coarsest, or an external justification of the semantic anchors. Point identification itself is exact by definition when $|\mathcal{K}_{\Pi,\Omega}| = 1$.

Proposition 2 (Componentwise sharpness for uniform recovery)**.** *The two-stage interface is componentwise sharp in the following sense. At extraction, omitting any one of H, P_R, Λ admits finite pairs with identical retained source fields but different E, R, or A touch. After touch extraction, uniformly erasing any one of terminal semantics (closed/open status together with the closed target class when defined), touch, cost, positive-support branching, or controller observation admits a finite pair that becomes indistinguishable under the weakened kernel while some closure/freeze value changes. Source-exposed atomic actions/checkpoints form the domain on which touch and freezing are defined rather than a detachable scalar field. The closed target class remains in the kernel for target-semantic fidelity; no separate componentwise-necessity claim is made for it apart from the terminal-semantics block.*

Proposition 3 (Rectangular partial identification)**.** *Suppose all $C \in \mathcal{C}(\Omega)$ share the same finite kernel except for independently variable R-touch membership. For an action set B, write $\kappa^{\neg B}_\Pi$ for the value after removing exactly the actions in B. Let Q^{def}_R touch R in every completion and Q^{poss}_R touch R in at least one. Then for every completion*

$$\kappa^{\neg Q^{\mathrm{def}}_R}_\Pi \le \kappa^{\neg\{R\}}_{\Pi,C} \le \kappa^{\neg Q^{\mathrm{poss}}_R}_\Pi .$$

If rectangularity admits the endpoint assignments, both bounds are sharp. With a common finite unrestricted value, the endpoints give the exact extended-real extrema/classification of $D_R(\Omega) = \{\Delta^C_R\}$.

4 GEOMETRY ONCE THE COUNTERFACTUAL IS IDENTIFIED

Table 1 reports R/baseline verdicts and boundary status; Proposition 6 separately captures strict complementarity.

Identified-set verdict	**Condition over all** $C \in \mathcal{C}(\Omega)$
Robust R-zero	every completion has finite unrestricted value and $\Delta_R^C = 0$
Robust finite-positive	every completion has finite unrestricted value and $0 < \Delta_R^C < \infty$
Robust structural	every completion has finite unrestricted value and $\Delta_R^C = \infty$
Robust baseline-unclosable	every completion has unrestricted value $\kappa_\Pi^C = \infty$
Boundary-underidentified	completions cross preceding verdicts, resource types, or K_Π signatures

Table 1: The first four rows are R/baseline verdicts; boundary underidentification is orthogonal and may coexist.

Proposition 4 (Zero-gap criterion)**.** *For a finite PC instance with nonnegative costs and attained finite unrestricted and R-frozen optima, $\Delta_R = 0$ iff some unrestricted optimum uses no R-touching action on any positive-support branch.*

This is intentionally elementary policy-class inclusion, not a new dynamic-programming principle.

Policy sufficiency. In Proposition 5, observation-equivalent checkpoints must agree on closed/open status, source action identities/touch, and successor observation classes, so each attractor layer admits an ω-measurable witness action. Appendix G gives the formal condition and proof.

Proposition 5 (Exact non-R barrier)**.** *On a finite policy-sufficient checkpoint graph, let W_0 be the closed checkpoints and*

$$W_{k+1} = W_k \cup \{s : \exists q \in Q(s),\ R \notin T_s(q),\ \mathrm{Succ}^+(s,q) \subseteq W_k\}.$$

At the finite fixed point $W_{\neg R}$, a finite sure-closing non-R policy exists exactly from states in $W_{\neg R}$; the complement is the exact structural representation barrier.

Proposition 6 (Strict evidence/representation complementarity)**.** *For every $d \geq 1$ there exists a finite PC family in which every closing branch must acquire all d independent evidence bits and execute all d corresponding representation refinements. Evidence-only and representation-only restrictions are each nonclosing, while their governed composition closes at cost $2d$. Authority is fixed throughout.*

Proposition 7 (Target refinement and exact transfer)**.** *Fix one PC system and change only the sealed target. Say T_2 refines T_1 on a represented fiber when $T_2(x) = T_2(x')$ implies $T_1(x) = T_1(x')$ for every x, x' in that fiber. If this holds on every reachable represented fiber, then $\kappa_{\Pi,T_1}^{\neg F} \leq \kappa_{\Pi,T_2}^{\neg F}$ for all $F \subseteq G$. If the targets induce the same equality partition on every reachable represented fiber, all coordinates agree and $K_{\Pi,T_1} = K_{\Pi,T_2}$.*

The complementarity result matters for interpretation: E and R are not interchangeable names for generic "information." A fact can be absent, admitted but aliased, or admitted and represented yet unusable under current authority.

5 Natural identification boundaries and controlled geometry

The empirical stack is layered rather than treated as one universal benchmark. The natural pair tests whether Stage-III identification can fail or succeed in independently engineered systems; PC-TAU then supplies controlled nondegenerate geometry, and LLM1 asks whether trained agents can realize the legal fallback.

5.1 Polaris: natural failure of resource identification

Apache Polaris PR #4992 predates PC and repaired an omitted realm identifier in the OPA authorization input used for tenant isolation Apache Polaris (2026). We define $\Omega_{\mathrm{Polaris}}$ from the pre-repair source, PR semantics, and public realm/OPA architecture—not code alone Apache Polaris Realm (2026); Apache Polaris OPA (2026). These artifacts fix that `RealmContext` contains the realm before OPA payload construction and that the patch threads it into `input.context.realm`, but do not define the PC admission predicate or certificate equivalence there. Both completions preserve those fixed facts: one treats the realm as already admitted and the call as an R refinement; the other admits it at OPA-message construction with fixed P_R, making the same transition E-touching. Hence

$$\boxed{\Omega_{\mathrm{Polaris}} \text{ is } E/R \text{ boundary-underidentified.}}$$

System	Identification result	Scientific job
Apache Polaris	Dataflow fixed; admission/certificate boundary not unique (E/R underidentified).	Natural negative: refuse unique attribution.
XACML/ AuthzForce	Measured interface fixed; pure E touch and singleton signature.	Natural positive Stage III.
PC-TAU Exact	Complete constructed interface with finite substitution plus all four exact regimes.	Controlled freeze geometry.
LLM1	Anchors inherited; nonzero observed $F010$ fallback discovery among resolved episodes.	Learned realizability.

Table 2: Empirical legs and evidentiary roles. None is presented as proof of a theorem.

If additional governing evidence fixed the omitted boundary, it would refine Ω and PC would recompute the identified set rather than preserve the ambiguity. The contract-relative computation and independent assurance remain valid but do not override this source-relative conclusion (Appendix J).

5.2 XACML/AUTHZFORCE: NATURAL POSITIVE STAGE III

The positive witness `PC-XACML-S3PLUS-v1` uses OASIS XACML 3.0 and AuthzForce Core CE 21.2.0, pinned at commit `3cc0e988e1639da48184434cd5c918102ff5b499`, with the pre-existing `StatusDetail.MissingAttributeDetail` fixture OASIS (2013); AuthzForce (2026). Its original request is missing a subject attribute required by a frozen-policy `MustBePresent=true` designator. Two completed worlds keep subject, resource, action, policy, configuration, engine, and every other request field fixed; only the missing attribute value differs. The sole unit-cost repair q_{supply} copies the request, supplies that world-specific attribute using the identifier/category/datatype already exposed by the PDP, and submits it to the same native engine. Native policy and engine are unchanged and the wrapper performs no authorization logic.

Source-entailment, not chosen typing. $\Omega^{\star}_{\mathrm{XACML}}$ fixes H to the native PDP-visible request context, records the required attribute as absent there with no PIP/selector admission channel, fixes P_R by the frozen policy designators, Λ by the fixed policy/provider/preprocessor/interface, and `Atom` by the PEP-style request/PDP/response transaction. Hence every compatible completion has $H(s') \neq H(s)$, $P_R(s') = P_R(s)$, $\Lambda(s') = \Lambda(s)$ for q_{supply}, so $T(q_{\mathrm{supply}}) = \{E\}$. Wrapper-local possession is not prior admission under this source-fixed H. Appendix I gives the source propositions and entailment proof; experiment-authored measurement conventions remain explicitly separated.

After authorization, the native PDP produced

$$x_{\mathrm{permit}} \mapsto \mathrm{Permit}, \qquad x_{\mathrm{nonpermit}} \mapsto \mathrm{NotApplicable},$$

matching the preregistered target classes. Dual conformance verifiers passed before outcome opening. The PC contract was then compiled with no resource-touch field; two independent touch derivations agreed on $T(q_{\mathrm{supply}}) = \{E\}$. Independent exact solvers returned

$$\boxed{K_{\Pi} = [1, \infty, 1, 1, \infty, \infty, 1, \infty]}, \qquad \boxed{|\mathcal{K}_{\Pi, \Omega^{\star}_{\mathrm{XACML}}}| = 1},$$

with a completion certificate containing zero free K-relevant fields. The classification is `STRUCTURAL_E_TOUCH_DEPENDENCE`. The final falsification phase included corruptions, semantic canaries, interface-preserving perturbations, explicit interface-change refusal, cost/negative-world sweeps, anchor ablations, manual-label injection, leakage checks, and clean reproduction. This is one source-native positive specification witness, not prevalence, production safety, universal XACML behavior, or causal decomposition inside mixed actions.

5.3 CONTROLLED GEOMETRY AND LIVE REALIZABILITY

PC-TAU is source-anchored constructed over pinned τ-family airline and retail tasks Yao et al. (2024). Across 135 primary instances, the pure R optimum has

$$K_{\Pi} = [1, 1, 2, 1, \infty, 1, \infty, \infty], \qquad \Delta_R = 1 :$$

freezing R removes the cost-1 action but leaves a legal two-step fallback. Its factorial panel separately realizes R-zero, finite-positive, structural, and complementarity regimes. Thus PC-TAU studies geometry once the interface is fully declared; it carries no natural-prevalence claim.

A preregistered live follow-on study scheduled 5,436 provider-verified cells over three model families; $F000/F010$ denote unrestricted/R-frozen conditions. The ledger records 1,982 terminal transport failures, 145 interaction-budget stops, and 3,309 resolved cells; transport failures are behavioral nulls under the frozen transport-only retry rule. Because the conditions resolve on different subsets, marginals are descriptive, not paired causal. Under $F010$, resolved fallback discovery was $.105/.130/.365$ for DeepSeek/Qwen/GLM. Hard-gate observations and the matched 120-decision semantic-repair motivation are retained in Appendix J.

6 Falsification, closest work, and claim boundaries

Falsification discipline. The historical program retains negative/null outcomes. XACML is fail-closed: external specimen/targets were frozen, failed semantic certifications preserved, touch mechanically derived, independent solvers and corruption/ablation checks used, and no post-outcome semantic repair permitted. Full counts and chronology are in Appendices H–J.

6.1 Closest objects and novelty boundary

The point is not that prior authorization analysis cannot evaluate changes. The separation is that such analyses generally condition on an already supplied policy/state/intervention semantics, whereas PC asks whether source evidence identifies the *resource counterfactual* being analyzed. Corollary 1 makes this formal.

Closest-work separation. Margrave Fisler et al. (2005) compares supplied XACML policy versions; dynamic/state-modifying authorization Dougherty et al. (2006); Becker and Nanz (2007) reasons over fixed state/update semantics; trust negotiation/decision-region methods Winsborough and Li (2006); Javdani et al. (2014) condition on disclosure/acquisition semantics; partial ID/causal abstraction Kline and Tamer (2023); Beckers and Halpern (2019); Li et al. (2025) provide generic identified-set/intervention machinery; and explainable-access-control causality Mehri et al. (2026) computes counterfactual causes from formalized policy features/effects. PC instead asks whether Ω identifies the $E/R/A$ intervention upstream; Corollary 1 gives a same-source counterexample. Recent agent-authorization work is neighboring problem-convergence evidence, not prevalence evidence (Appendix J).

Novelty boundary. PC does not claim novelty for reachability, dynamic authorization, information acquisition, generic partial identification, bisimulation, policy differencing, or generic observation/intervention separation. Its claim is the authorization-resource intervention identification problem itself:

$$\boxed{\Omega \longmapsto \mathcal{C}(\Omega) \longmapsto \mathcal{T}_\Omega \longmapsto \mathcal{K}_{\Pi,\Omega}},$$

plus the behavioral/source non-identifiability results, conditional anchored recovery, and matched operational freezes.

Scope and limitations. $E/R/A$ is declared, not universal; Stage III is conditional; K_Π is source-interface-relative; Δ_R measures action availability, not internal causal contribution; and resource-support bisimulation is sufficient only. Polaris is a natural negative, XACML one positive specification witness, PC-TAU constructed, LLM1 descriptive, and the hard gate is not a safety theorem. Appendix J retains the full ten-item boundary.

7 Conclusion

Perceptive Closure separates identification from resource-intervention geometry. Theorems 1–2 show behavior/dataflow need not identify the counterfactual; Theorem 3 gives conditional recovery. Polaris and XACML/AuthzForce realize both sides naturally; PC-TAU and LLM1 study downstream geometry and realizability. Planning evaluates a chosen intervention; Perceptive Closure asks whether it is identified.

Reproducibility statement

The paper-facing experiments are sealed separately with frozen source/provenance manifests, exact contracts or schedules, independent recomputation, corruption/falsification checks, and deterministic replay where applicable. Complete proofs and theorem-adjacent finite checks are in the appendix. The XACML package additionally preserves the pinned external specimen and engine/policy/configuration, preregistered worlds/targets/freezes, failed certifications, final launch packet and conformance registry, outcome capsules, mechanical touch/freeze derivations, completion certificate, final audit, and deterministic archive; historical negative seals are preserved rather than rewritten.

AI use statement

Generative AI tools were used to help develop and critique conceptual/theoretical framing, formulate and stress-test mathematical claims, assist with proof drafting and counterexample search, triage literature, support implementation/code work, audit consistency, and draft/edit manuscript text. The authors reviewed the mathematical arguments, source evidence, code, experimental gates, and final wording and take responsibility for all claims and citations.

Ethics statement

The security-relevant natural examples use public, already disclosed artifacts and do not report a new exploitable vulnerability. Polaris PR #4992 is a public repaired issue; PC uses it only to study semantic identification of the resource counterfactual. The paper makes no deployment-safety claim and preserves negative/failure evidence rather than suppressing it.

References

Apache Polaris. *fix(authz): include realm identifier in OPA authorization input for tenant isolation*. Pull request #4992, 2026.

Apache Polaris. *Realm*. Public documentation describing `RealmContext`, multitenant isolation, and authorization context, 2026.

Apache Polaris. *Open Policy Agent (OPA) Integration*. Public documentation of the external authorization interface, 2026.

AuthzForce. *AuthzForce Core CE 21.2.0*, release-21.2.0, commit `3cc0e988e1639da48184434cd5c918102ff5b499`, 2026.

Moritz Y. Becker and Sebastian Nanz. A logic for state-modifying authorization policies. *ESORICS 2007*, LNCS 4734, 203–218, 2007. DOI: 10.1007/978-3-540-74835-9_14.

Sander Beckers and Joseph Y. Halpern. Abstracting causal models. *AAAI* 33(01), 2678–2685, 2019.

Daniel J. Dougherty, Kathi Fisler, and Shriram Krishnamurthi. Specifying and reasoning about dynamic access-control policies. *IJCAR*, 632–646, 2006.

Kathi Fisler, Shriram Krishnamurthi, Leo A. Meyerovich, and Michael Carl Tschantz. Verification and change-impact analysis of access-control policies. *ICSE*, 196–205, 2005. DOI: 10.1109/ICSE.2005.1553562.

Zijie Huang. Self-Certification of Representation Adequacy: Sequential Certification at Minimum Task Loss. arXiv:2608.02267, 2026.

Shervin Javdani, Yuxin Chen, Amin Karbasi, Andreas Krause, Drew Bagnell, and Siddhartha Srinivasa. Near optimal Bayesian active learning for decision making. *AISTATS*, 430–438, 2014.

Brendan Kline and Elie Tamer. Recent developments in partial identification. *Annual Review of Economics* 15, 125–150, 2023.

Xiusi Li, Sekou-Oumar Kaba, and Siamak Ravanbakhsh. On the identifiability of causal abstractions. *AISTATS*, PMLR 258:3241–3249, 2025.

Gelareh Hasel Mehri, Clemens Dubslaff, Tim A. C. Willemse, and Nicola Zannone. A causal framework for explainable access control. *ACM SACMAT*, 149–155, 2026. DOI: 10.1145/3750555.3811906.

OASIS. *eXtensible Access Control Markup Language (XACML) Version 3.0*. OASIS Standard, 2013.

Jaehong Park and Ravi Sandhu. The UCONABC usage control model. *ACM TISSEC* 7(1), 128–174, 2004.

Krti Tallam. Partial Evidence Bench: Benchmarking Authorization-Limited Evidence in Agentic Systems. arXiv:2605.05379, 2026.

William H. Winsborough and Ninghui Li. Safety in automated trust negotiation. *ACM TISSEC* 9(3), 352–390, 2006.

Zheng Yan, Jingxiang Weng, Charles Chen, et al. Do Coding Agents Understand Least-Privilege Authorization? arXiv:2605.14859, 2026.

Shunyu Yao, Noah Shinn, Pedram Razavi, and Karthik Narasimhan. τ-bench: A benchmark for tool-agent-user interaction in real-world domains. arXiv:2406.12045, 2024.

Qiuyang Zhan, Rui Zhang, Sheng Guo, Lepeng Zhao, and Zhuotao Liu. When Memory Becomes Authority: Benchmarking Authority Collapse at the Memory Consolidation Boundary. arXiv:2608.01679, 2026.

Yifan Zhang, Xinkui Zhao, Sai Liu, Hengxuan Lou, Guanjie Cheng, and Chang Liu. FAVA: Formal Authorization for Verified Agents with Evidence-Backed Permission Graphs. arXiv:2607.27267, 2026.

Reshabh K. Sharma, Linxi Jiang, Zhiqiang Lin, and Shuo Chen. PAuth – Precise Task-Scoped Authorization For Agents. arXiv:2603.17170, 2026.

Shidong Pan, Xiaoyu Sun, Tianyi Zhang, Dianshu Liao, Meixue Si, and Zhenchang Xing. SkillGuard: A Permission Framework for Agent Skills. arXiv:2606.03024, 2026.

Wujie Xiong, Rabimba Karanjai, Yang Lu, Weidong Shi, and Lei Xu. Reachability-Based Capability Confinement for LLM Agents under Indirect Prompt Injection. arXiv:2608.30041, 2026.

Igor Santos-Grueiro. Temporary Authority, Permanent Effects: Commit-Time Authorization for LLM Agents. arXiv:2607.10487, 2026.

James C. Davis, Kelechi Kalu, Huiyun Peng, and Parth V. Patil. Model-Based Agentic Software Engineering. arXiv:2608.25174, 2026.

APPENDIX TYPE LEDGER: FORMAL CONTRACT TYPES AND WELL-FORMEDNESS

For completeness, let $\mathcal{Q}$ be the global source-action identity set and $\mathcal{Y}$ the controller-visible semantic outcome-class set. Elements of U_H are canonical normalized-history representatives. The value $H(s)$ indexes the current admitted-history certificate cell; source-compatible histories need only occur in $[H(s)]_{P_R(s)}$, and source compatibility is carried explicitly by $\sim$. Thus a canonical representative used to name a cell need not itself be the world-specific compatible history. The contract components have types

$$\sim \subseteq X \times U_H, \quad s_0 \in S, \quad H : S \to U_H, \quad P_R : S \to \mathrm{Eq}(U_H), \quad \Lambda : S \to \mathfrak{A}, \quad \omega : S \to O,$$
$$Q : S \to 2^{\mathcal{Q}}, \quad \mathrm{Succ}^+ : \{(s,q) : q \in Q(s)\} \to 2^S \setminus \{\varnothing\},$$
$$Y : \{(s,q,s') : q \in Q(s),\ s' \in \mathrm{Succ}^+(s,q)\} \to \mathcal{Y}, \quad c : \mathcal{Q} \to \mathbb{R}_{\geq 0}, \quad A_\Pi : X \to D.$$

Atom is the declared partition of source-level transition events into the atomic action identities in $\mathcal{Q}$ together with their pre/post checkpoint locators. It must agree with Q and Succ^+. For every reachable s, $P_R(s)$ is an equivalence relation on U_H and $C_R(s) \neq \varnothing$; Y is total on positive-support edges; costs are finite and nonnegative; and the authority-coherence condition in Section 2 holds. A closed target label is the unique element of $\{A_\Pi(x) : x \in C_R(s)\}$ and is defined only when s is closed. These conditions make the closure, touch, policy, kernel, and source-completion objects used below well-typed.

A Proof of Proposition 1: control–certificate separation

Take two source-valid histories h_0, h_1 generated in worlds x_0, x_1 with opposite target authorizations, and two source-reachable open checkpoints u_0, u_1 with $H(u_i) = h_i$. Let the same certificate relation P_R place h_0, h_1 in one cell at both checkpoints, so both remain target-open, while the controller observation distinguishes them: $\omega(u_i) = i$. Both unit-cost actions q_0, q_1 are legal. At u_i, action q_i returns an admitted world-specific witness whose successor certificate cell is target-homogeneous; q_{1-i} returns a common null outcome and exhausts budget $B = 1$ without closing. Hence the ω-measurable controller choosing q_i at observation i closes from each source-reachable checkpoint within budget one.

Suppose one quotient Z on histories is used both for controller measurability and terminal certification. If $Z(h_0) = Z(h_1)$, a Z-measurable controller must use the same action distribution at u_0, u_1; under the worst-support criterion, one reachable case receives the wrong action with positive support and remains open at exhausted budget, so Z is not control-complete. If $Z(h_0) \neq Z(h_1)$, using Z as the certificate partition separates opposite target worlds before admitted repair evidence arrives, incorrectly certifying the pre-repair checkpoints. Therefore no single untyped quotient satisfies both requirements. □

B Proof of Theorem 1: behavioral non-identifiability

Fix $g > 0$ and worlds x_0, x_1 with $A_\Pi(x_0) \neq A_\Pi(x_1)$. Both systems have one open initial checkpoint s and two source actions: fast q of cost 1 and backup b of cost $1 + g$. In world x_i, q yields controller-visible outcome y_i and a closed terminal t_i^q; b yields z_i and a closed terminal t_i^b. Action identities, controller observations, visible outcomes, costs, positive support, terminal target labels, and stopping are identical across systems.

In M'_g, choose a fixed certificate partition that already distinguishes both y- and z-extended admitted histories; both actions only change admitted history, so $T(q) = T(b) = \{E\}$. In M_g, choose a pre-q partition that aliases the two y histories and a post-q partition that separates them, while the z histories are already separated by the unchanged partition. Thus q atomically touches $E + R$ and b touches E only. If realized representation values are controller-visible, rename values so the visible pre/post sequence is identical; partition equality, not value names, defines R touch.

Every unrestricted policy therefore induces the same labeled support tree, so $M_g \equiv_{\mathrm{ctrl}} M'_g$. In both systems q closes at cost 1. Under an R freeze, q remains available in M'_g and the value remains 1; in M_g, q is removed and the cheapest closer is b at $1 + g$. Thus $\Delta_R = g$ versus 0. Deleting b from both systems preserves unrestricted equivalence and value 1, while the R freeze leaves M'_g closable by q and M_g with no closer, giving 1 versus ∞. □

C Proof of Theorem 2: source-semantic boundary non-identifiability

Fix $g > 0$. The implementation description Ω_g contains worlds x_0, x_1 with opposite target classes, one open entry point, a raw context field $d_i = i$ already present in implementation memory in world x_i, a fast source action q of cost 1 that copies that field into the authorization-interface payload and returns a target-closed terminal, and a backup action b of cost $1 + g$ that obtains a separate signed witness and also closes. Ω_g fixes these raw dataflows, action identities, costs, visible outcomes, terminal target labels, and the implementation transition graph. It deliberately does not specify whether the pre-call context field is already part of normalized authorization-admitted history, nor the pre-call certificate equivalence relation.

Construct C_R as follows. Let the single initial representative be $H(s_0) = h_\star$, and let its certificate cell $[h_\star]_{P_R(s_0)}$ contain source-compatible histories h_0, h_1 carrying d_0, d_1. The two source-visible outcome branches of q keep H and Λ fixed while refining P_R so the successor represented fiber is target-homogeneous; therefore $T^{C_R}(q) = \{R\}$. The backup b instead moves H to an outcome-specific signed-witness representative under an unchanged certificate relation, so b is E-only.

Construct C_E on the same implementation. Let $H(s_0) = u_\star$ represent the common pre-admission history, which omits the raw field d_i even though that field remains in implementation memory as required by Ω_g. Use one fixed certificate relation whose pre-admission cell is target-heterogeneous but whose outcome-specific post-admission representatives h_i occupy target-homogeneous cells. Action q moves H from $u_\star$ to the corresponding h_i while leaving P_R and Λ fixed, so $T^{C_E}(q) = \{E\}$. Define b as the same E-only signed-witness path. Both contracts project to every field constrained by Ω_g; they differ only on the semantic admission/certificate anchor that Ω_g omitted. Their controller-visible action/outcome/support trees are identical under every unrestricted policy.

The fast action closes at cost 1 in both. Under an R freeze, q is available in C_E and the value remains 1; in C_R, q is removed and b gives value $1 + g$. Removing b produces 1 versus ∞. Hence one source description can admit complete E- and R-typed contracts with different resource counterfactuals. □

D Proof of Theorem 3: anchored identification and invariance

Part (i): touch identification. Once H, P_R, Λ are fixed extensionally on every source-exposed positive-support successor, Equation (1) is a deterministic set-valued function of their pre/post equality relations. Source-exposed atomicity fixes which semantic effects belong to one intervention rather than to independently selectable subactions. Therefore every touch set is unique; no manual $E/R/A$ label remains free.

Part (ii): kernel transport. Fix a freeze $F \subseteq G$, an interface correspondence (α, β), and an interface-equivariant resource-support bisimulation B between initial checkpoints s_0, s_0'. Because α_o preserves touch, an action survives the F freeze on one side iff its mate survives on the other. We prove by induction on the height of a proper closing policy tree that any such tree rooted at s transports to a tree of the same worst-branch cost rooted at every related s'.

At height zero, s is closed; B preserves terminal/open status and target class, so s' is closed with the same zero remaining cost. For height $h + 1$, let the policy choose surviving action q at s with $o = \omega(s)$. The correspondence supplies $q' = \alpha_o(q)$ with identical cost/touch and $\beta_{o,q}$ bijects its positive-support outcome classes. Each successor pair lies in B, so induction transports every subtree with the same worst-branch cost. Taking the maximum over the bijected outcomes and adding the common action cost preserves the tree value. Crucially, α_o depends only on the controller observation: two aliased checkpoints using the same o receive the same action renaming, so transported choices remain ω-measurable. Thus every proper closer of value v transports to one of value v; symmetry gives the converse. This holds for every freeze F, proving coordinatewise equality of K_Π and all derived finite/structural classifications.

The argument does not require optimal attainment. If one side has no proper closer under a freeze, a finite closer on the other would transport back and contradict that fact; therefore both values are ∞.

Part (iii): point identification. If every completion in $\mathcal{C}(\Omega)$ belongs to one interface-equivariant resource-support kernel class, Part (ii) assigns the same K_Π to every completion, so $\mathcal{K}_{\Pi,\Omega}$ is a singleton. A source refactoring may therefore rename actions/outcomes and change implementation-local syntax/internal plumbing, provided the observation-consistent interface correspondence and anchored semantic kernel are preserved; such a refactoring cannot change the signature. □

E Proof of Proposition 2: componentwise sharpness

The extraction and planning layers require different witnesses.

Extraction layer. (H) Take the same source transition with fixed P_R, Λ; in one completion its normalized admitted history changes and in the other it does not. Omitting H makes the retained source fields identical while E touch differs. (R) Hold H, Λ fixed; in one completion the post-state uses a strictly finer certificate partition, while in the other only a raw representation value is renamed and the equivalence relation is unchanged. Omitting P_R loses R touch. (A) Hold H, P_R and currently enabled actions fixed, but let the post-state authority operator differ on a future source-valid interface in one completion and remain extensionally equal in the other. Omitting Λ loses A touch. Each pair is finite.

Planning layer. After touch has been extracted, the following finite pairs prove that each listed kernel field is necessary for uniform preservation.

Erased field	Pair made indistinguishable by erasure	Changed PC quantity
Terminal semantics	Two actionless checkpoints; one closed with its target label, one open; erasing the status-plus-dependent-label block makes them indistinguishable.	$\kappa : 0$ vs ∞
Touch	Same unit-cost closing action and visible outcome; touch $\{E\}$ vs $\{E, R\}$.	frozen-R: 1 vs ∞
Cost	Same one-action closer with cost 1 vs 2.	$\kappa : 1$ vs 2
Positive support	Same visible outcome; one action supports only a closed successor, its mate also supports an open absorbing trap.	finite vs ∞
Controller observation	First action branches to u_0, u_1; only a_i closes from u_i. One system reveals i, the other aliases them under ω.	finite vs ∞

For the observation row, any ω-measurable policy in the aliased system must choose the same continuation distribution at u_0, u_1; worst positive support therefore includes a wrong continuation and no sure closer exists. The first row erases terminal semantics as one dependent block because a closed target label is defined only for a closed checkpoint; the kernel still retains that label for target-semantic transport, but Proposition 2 makes no separate necessity claim for it. □

F Proof of Proposition 3: rectangular partial identification

For any completion C, rectangular ambiguity implies

$$Q_R^{\mathrm{def}} \subseteq Q_R(C) \subseteq Q_R^{\mathrm{poss}},$$

where $Q_R(C)$ is the set removed by an R freeze. Removing more actions can only shrink the feasible proper-policy set, so closure cost is monotone nondecreasing under action removal. Hence

$$\kappa_{\Pi}^{\neg Q_R^{\mathrm{def}}} \leq \kappa_{\Pi,C}^{\neg\{R\}} \leq \kappa_{\Pi}^{\neg Q_R^{\mathrm{poss}}}.$$

If the ambiguity set is rectangular, one completion assigns R exactly to the definite set and another assigns it to the possible set; these attain the endpoints. If unrestricted κ_Π is common and finite, subtracting it in the extended reals gives the sharp $D_R(\Omega)$ extrema. If transitions, costs, support, checkpoints, or authority dynamics also vary across completions, this two-solve formula is not claimed. □

G Proofs of Propositions 4–7

Zero-gap criterion. Let $\mathcal{R}$ be the unrestricted proper-policy class and $\mathcal{R}_{\neg R} \subseteq \mathcal{R}$ the subclass containing no R-touch action on any positive-support branch. By action removal, $\kappa_\Pi^{\neg R} \geq \kappa_\Pi$. If an attained unrestricted optimum ρ^* lies in $\mathcal{R}_{\neg R}$, then the frozen value is at most its cost κ_Π, hence equality. Conversely, if $\Delta_R = 0$, the attained finite frozen optimum is also an unrestricted policy of value κ_Π and lies in $\mathcal{R}_{\neg R}$ by construction. □

Exact non-R barrier. Formally, policy sufficiency requires that $\omega(s) = \omega(s')$ imply: (i) equal closed/open status; (ii) the same source action identities; and (iii) for each common action q, equal touch and equal successor-observation sets, $T_s(q) = T_{s'}(q)$ and $\{\omega(u) : u \in \mathrm{Succ}^+(s, q)\} = \{\omega(u') : u' \in \mathrm{Succ}^+(s', q)\}$. Because the checkpoint set is finite, the increasing sequence W_k reaches a fixed point. First show by induction that every W_k is a union of controller-observation classes. It is true for W_0 by policy sufficiency (equal observations preserve closed/open status). Suppose it holds for W_k and let $s \in W_{k+1} \setminus W_k$ with witness action q. For any s' with $\omega(s') = \omega(s)$, policy sufficiency exposes the same action identity q, preserves its non-R touch, and gives the same successor observation classes. Since W_k is an observation-class union, every positive-support successor of (s', q) is also in W_k, so $s' \in W_{k+1}$. Hence the same source action identity is a valid ω-measurable choice throughout each newly added observation class. Induction on k therefore

constructs a finite sure-closing non-R policy for every state in the fixed point. Conversely, if $s \notin W_{\neg R}$, every non-R action has a positive-support successor outside $W_{\neg R}$. Choosing such a successor keeps a branch outside closure (or reaches an open state with no surviving action), contradicting finite sure closure. Thus the complement is exactly the structural non-R barrier under the stated policy-sufficiency condition. □

Strict evidence/representation complementarity. For depth d, let worlds be $x \in \{0,1\}^d$ and target $A_\Pi(x) = \bigoplus_{j=1}^{d} x_j$. The semantic state contains an admitted-coordinate set I and certificate-exposure set J, initially empty. Evidence action e_j adds j to I and records x_j without changing J or authority. Representation action r_j adds j to J, refining the certificate partition so histories are distinguished by coordinate j whenever that coordinate is admitted; it does not change I or authority. Thus e_j is pure E and r_j pure R, regardless of order. If $j \notin I$, two worlds differing only in bit j remain compatible. If $j \in I$ but $j \notin J$, the certificate still aliases histories differing only in that admitted-but-hidden bit. In either case opposite parity targets remain compatible. Hence every closing branch uses all d evidence actions and all d representation actions. Executing each pair once gives $I = J = \{1, \dots, d\}$ at cost $2d$. □

Target refinement and exact transfer. Fix a freeze F. If T_2 refines T_1, every represented fiber homogeneous for T_2 is homogeneous for T_1. Therefore every proper T_2-closing policy under the freeze is also a proper T_1-closing policy with identical branch costs; minimizing gives $\kappa^{\neg F}_{\Pi,T_1} \le \kappa^{\neg F}_{\Pi,T_2}$. If the two targets induce the same equality partition on every reachable represented fiber, open/closed status agrees at every checkpoint. Actions, freezes, costs, support, and observations are unchanged, so every policy has the same closure status and cost for both targets; taking infima gives coordinatewise equality of K_Π. □

H Executable and empirical theorem-adjacent checks

These checks are not proofs; they are implementation realizations of already proved objects.

Same-instance structural witness. One immutable serialized PC instance holds worlds, target, evidence, authority, controller, actions, costs, outcomes, and transitions fixed; the frozen evaluator only filters R-touch actions. Exhaustive uniform-cost search returns unrestricted 1 and frozen-R ∞ with zero newly acquired evidence; an independently structured FIFO oracle agrees.

PC-TAU exact audit. The sealed exact layer contains 1,336 rows: 1,080 primary and 256 factorial. An independent Phase-4 planner re-solves all 1,080 primary cells with 100% agreement; same-instance hashes prove only the freeze mask varies across each eight-cell group. The post-measurement audit narrows the interpretation to the artifact's source-anchored constructed class SG-B: one shared route topology across three mechanism families.

LLM1 audit. The 5,436 scheduled keys are complete and unique after disclosed infrastructure remediation. Independent reconstruction matches all 5,436 metric rows, verifies 1,359 pairing quartets, provider locks, token caps, costs, and terminal classes, and all 24 deliberate corruption families fail closed. A terminal transport failure means the frozen transport-only retry rule ended without a usable provider trajectory; the episode history includes 1,982 such failures and 145 interaction-budget stops. Both remain null behavioral observations rather than being imputed as agent failures.

Independent theorem-sanity program. A separate finite checker reproduces the Stage-I finite-gap/structural witnesses and Stage-II E-versus-R value witness; preserves all eight freeze values under 250 random kernel refactorings; matches the non-R attractor against exhaustive policy enumeration on 120 random fully observed games; verifies rectangular partial-identification bounds over 160 random kernels; checks target-refinement monotonicity over 180 random systems; and exhaustively verifies the parity complementarity invariant for $d = 1, \dots, 6$. These are theorem-adjacent implementation checks, not substitutes for the proofs or proof-assistant verification.

I XACML audit chronology and non-regression disclosure

Source propositions and entailment. The Stage-III claim is not inferred from a preregistered PC encoding or from auditor agreement alone. Write $\Omega^\star_{\mathrm{XACML}}$ for the source/specification evidence

consisting of the following externally grounded propositions: (X1) H is the canonicalized logical XACML request-context attribute multimap available to native PDP evaluation, and the required world-specific attribute is absent from the original request; (X2) the pinned fixture configures no PIP source or selector/XPath channel that supplies that attribute before q_{supply}; (X3) P_R is induced by the frozen policy's reachable `AttributeDesignator` coordinates under normative retrieval/matching semantics and remains fixed during the repair; (X4) policy, provider, preprocessor, request interface, and native PDP remain fixed, so Λ is extensionally unchanged; and (X5) the source-atomic repair copies the frozen request, inserts exactly the one missing attribute, and executes the same PEP-style request/PDP/response transaction with no wrapper authorization logic. Source classes are N1 = normative XACML standard, N2 = pinned external engine configuration/schema, N3 = pinned pre-existing fixture/policy/request/expected response, N4 = implementation-source corroboration, and X = experiment-authored orchestration/measurement.

By X1–X2, every $C \in \mathcal{C}(\Omega^{\star}_{\mathrm{XACML}})$ has $H(s') \neq H(s)$: possession of the value in wrapper-local world state is not admission into the source-fixed PDP-visible request context. By X3, $P_R(s') = P_R(s)$; by X4, $\Lambda(s') = \Lambda(s)$; and X5 fixes the intervention/atomicity boundary. Equation (1) therefore gives $T(q_{\mathrm{supply}}) = \{E\}$ in every compatible completion. With the same source-fixed successor structure and the preregistered measurement conventions, the exact freeze computation gives the common vector $[1, \infty, 1, 1, \infty, \infty, 1, \infty]$, hence $|\mathcal{K}_{\Pi,\Omega^{\star}_{\mathrm{XACML}}}| = 1$. The completion-space certificate's zero free K-relevant fields is audit corroboration of this implication, not its logical premise.

External-versus-experiment provenance ledger. The accepted anchors use H from request-context semantics plus the frozen request/processing path (N1/N3, corroborated by N4), P_R from reachable frozen-policy `AttributeDesignator`s and normative retrieval semantics (N1/N3), Λ from the frozen root policy, provider/configuration, request preprocessor, and native request interface (N1/N2/N3), and `Atom` from the externally meaningful PEP-style transaction (N1, corroborated by N4). The two latent worlds, unit cost, freeze order, exact solver conventions, and recording format are X and are reported as such; no universal XACML or universal PC-ontology claim follows.

The XACML result is retained with its pre-measurement hardening history rather than presented as if the final interface was obvious from the start. The external specimen, worlds, targets, action identity, cost, freeze order, expected touch/signature, and positive/negative outcome taxonomy were frozen before target execution. Early semantic certification exposed unresolved $H/P_R/\Lambda/\mathrm{Atom}$ support and was treated as apparatus-hardening input while target execution remained zero. Subsequent hardening preserved failed panels and negative seals.

The operative launch architecture separated (i) static semantic facts that had to be fixed before execution from (ii) run-conformance facts whose truth necessarily depended on the future invocation but whose predicates, producers, and verifiers could be frozen in advance. Launch-003 used packet SHA-256 `8e23901d4643a42a4748af8894c0439f19f071f736d23225ea906aac97099ce6`, prompt SHA-256 `251f7e5875b243a24d0fe95b3252ecfdf08e06f97fbc254e0ed45ea3cb4c7865`, and registry SHA-256 `d7641f64ced4092a23e7e75dcb26f00f5fa274f27ca3171b2d39792e6fe7e44a`. Fresh launch auditors ultimately certified $H, P_R, \Lambda, \mathrm{Atom}$ as fixed and all six run-conformance obligations as legitimately deferred. Two intermediate AUD-L08 ingest attempts were procedurally invalid (first missing required packet/prompt hash bindings, then cross-reading other auditor identifiers) but raised no semantic objection; they were not recoded as scientific passes. No refreeze was triggered by those procedural errors.

After launch authorization, the first attempted x_{permit} run failed before any response/lock because the quarantine directory was absent. The chronology record was cleaned, the non-semantic directory bug was fixed, and the authorized run was retried; no outcome had been produced or opened. The completed target executions then produced hash-only capsules until the frozen dual conformance gate passed. An outcome-blindness verifier representation was corrected to use opaque path-SHA keys rather than fixture names containing decision-like substrings; derivation and equality checks were unchanged and the change was documented before outcome opening. Only after `RUN_CONFORMANCE_PASSED` were decisions parsed, followed by mechanical touch and freeze computation.

The final result is `POSITIVE_NATIVE_STAGE3`: $T = \{E\}$, $K = [1, \infty, 1, 1, \infty, \infty, 1, \infty]$, singleton identified set, no manual touch labels, and unchanged native policy/engine. The final audit

retains historical C01–C03 and C04–C06 failures, launch failures L01–L06, the AUD-L08 procedural invalid attempts, the pre-response quarantine-directory failure, the verifier-opacity correction, and the authorized-state skips. This disclosure is intentionally uglier than a clean rerun: the failed paths are part of the evidentiary record.

J Extended empirical details and complete claim boundaries

Motivating semantic repair. In a matched 120-decision semantic-repair study (artifact label DS-CF), predictive specialists were not retrained: observed unsafe acceptance remained 0 in both conditions while false rejection changed from 31.58% to 0%; ambiguous cases were escalated rather than forced into a binary decision. This motivates the distinction between predictive support and authorization representation; it is not evidence for the PC theorems.

Polaris admissibility ledger. $\Omega_{\mathrm{Polaris}}$ fixes the historical source/dataflow, the fact that `RealmContext` already contains the realm before OPA payload construction, PR #4992's threading of that realm into `input.context.realm`, and the public architectural facts that realm identifies the security domain and OPA consumes an authorization input. The checked governing artifacts do not state whether the preexisting realm is already part of normalized authorization-admitted history before serialization, nor do they define a PC certificate-equivalence relation at that boundary. The C_E and C_R completions are required to preserve every fixed operational fact and differ only on those omitted semantic anchors. This is a bounded admissibility claim, not a claim that no conceivable additional documentation could refine Ω; if such governing evidence exists, PC requires adding it and recomputing $\mathcal{K}_{\Pi,\Omega}$.

Polaris assurance. The contract-relative layer retains 26/26 source checks, 8/8 source-selected cases, 18/18 semantic locators, 12/12 contract comparisons, 1/1 independent touch reconstruction, 64/64 freeze cells, 8/8 case classifications, 10/10 controls, 12/12 corruption detections, and an 88-test byte-identical clean reproduction. These validate the contract-relative computation; they do not collapse source-relative E/R ambiguity.

PC-TAU exact audit. The primary population comprises 135 task instances, one shared route template, three mechanism families, five tool instances, and ten target-goal classes, classified in the artifact as source-anchored constructed SG-B (one shared topology across the three mechanism families). The exact seal contains 1,336 rows: 1,080 primary and 256 factorial. The factorial panel contains 8 R-zero, 8 finite-positive, and 16 structural cases, including 8 complementarity-flagged instances. An independent planner re-solves all 1,080 primary cells with 100% agreement.

LLM1 exact descriptive results. With $F000$ denoting unrestricted execution and $F010$ the R freeze, the primary resolved $n(F000)/n(F010)$ counts are $95/86$ (DeepSeek), $165/169$ (Qwen), and $229/178$ (GLM); the corresponding governance rates are $.179/.105$, $.370/.130$, and $.568/.365$. The run records 262 open-state submit attempts across 211 episodes; all were blocked and zero unauthorized effects executed. No causal freeze effect, population prevalence, budget-stop-as-failure sensitivity, or general safety claim is made.

Historical persistence. A 39-condition single-factor matrix retains expected pathologies, nulls, and non-degradation outcomes. A separate historical audit/falsification program, not one of the four paper-facing empirical legs, spans three generations and contains 90,000 canonical opportunities and 6.21M condition-level traces; independent recomputation checks 6,210,384 terminal records and 33,606,144 component comparisons with zero mismatch. These numbers support persistence and auditability, not measurements of Δ_R or independent population samples.

Complete limitations. (1) $E/R/A$ is a declared, componentwise-irredundant interface factorization, not a universal ontology. (2) Stage III is conditional on semantic anchors fixed by the governing specification/audited contract; PC does not decide where an omitted admission boundary "really" belongs. (3) K_Π is a property of the declared source action/checkpoint interface; interface-changing split/merge, batching, transactional decomposition, or checkpoint moves may change it. (4) Δ_R is an action-availability penalty, not causal decomposition inside a mixed action; reachable extensional side effects can induce touch even when not minimally target-causal. (5) Resource-support bisimulation is sufficient, not necessary/coarsest; the general nonrectangular completion problem is not algorithmically characterized, and no production-scale exact-solver complexity result is claimed. (6) Polaris naturally supports boundary underidentification; its structural-R value is contract-relative. (7)

XACML/AuthzForce supplies one natural positive specification witness, not prevalence or universal XACML behavior. (8) PC-TAU's 135 tasks share a constructed mechanism template and imply no natural prevalence. (9) LLM1's transport failures and budget stops have different scientific status; resolved marginals are descriptive and the claim is fallback realizability. (10) Zero unauthorized execution is a property of the evaluated hard gate, not a deployment-safety theorem.

J.1 Full closest-object comparison

Prior object	**Conditions on / answers**	**PC separation**
Margrave change impact Fisler et al. (2005)	Supplied XACML policy versions and environment constraints; characterize where decisions change.	Does Ω identify the $E/R/A$ intervention before its freeze consequence is evaluated?
Dynamic/state-modifying authorization Dougherty et al. (2006); Becker and Nanz (2007)	Stateful policy behavior, equivalence, and goal-directed authorization change.	Equivalent behavior or solvable repair can coexist with different resource-freeze dependence (T1/T2).
Trust negotiation / decision-region acquisition Winsborough and Li (2006); Javdani et al. (2014)	Disclosure or information acquisition until a decision/goal condition is reached.	PC separates admission, representation, and authority, then asks whether those intervention coordinates are identified.
Partial ID / causal abstraction Kline and Tamer (2023); Beckers and Halpern (2019); Li et al. (2025)	Generic identified sets, interventions, and abstraction/equivalence.	PC instantiates a source-relative authorization-resource object and exact freeze geometry; generic machinery is not claimed as new.
Explainable access-control causality Mehri et al. (2026)	Formalized policy features/effects; identify counterfactual feature causes and responsibility for authorization outcomes.	Conditions on an intervention vocabulary; PC asks whether source/specification evidence identifies the resource intervention coordinates upstream.
Agent authorization work Huang (2026); Tallam (2026); Yan et al. (2026); Zhan et al. (2026); Zhang et al. (2026); Sharma et al. (2026); Pan et al. (2026); Xiong et al. (2026); Santos-Grueiro (2026); Davis et al. (2026)	Separately studies evidence, representation, least privilege, memory/commit authority, permissions, and confinement.	Problem-convergence evidence for the dimensions PC asks to identify; no prevalence inference follows.

Table 3: Full closest-object comparison used for the novelty boundary.