

Information Barriers in Classification Systems: Matroid Structure and Witness Complexity

Tristan Simas

Abstract—We study the *identification capacity* of classification systems—the ability to determine which class an entity belongs to given limited observations. An observer receives binary query responses (“does entity v satisfy attribute I ?”) and must identify the entity’s class. We prove fundamental limits on identification cost as a function of observation model.

Information barrier. When distinct classes share identical attribute profiles, no algorithm—regardless of computational power—can identify the class from attribute queries alone. This is a channel capacity result: the mutual information $I(C; \pi(V)) < H(C)$ when the attribute mapping π is not injective on classes.

Rate-identification tradeoff. We analyze the tradeoff between *tag rate* L (bits per entity), *identification cost* W (queries per identification), and *error rate* D . A nominal tag achieving $L = \lceil \log_2 k \rceil$ bits for k classes yields $W = O(1)$ with $D = 0$. Without tags ($L = 0$), identification requires $W = \Omega(n)$ queries and may incur $D > 0$.

Converse. Any scheme achieving $D = 0$ requires $L \geq \log_2 k$ bits. This is tight: nominal tagging achieves the bound with $W = O(1)$.

Matroid structure. Minimal sufficient query sets form the bases of a matroid. The *identification dimension*—common cardinality of all minimal query sets—is well-defined and computable.

Applications. The theory instantiates to type systems (duck vs. nominal typing), databases (attribute vs. key lookup), and biological taxonomy (phenotype vs. species identifier). The unbounded gap $\Omega(n)$ vs. $O(1)$ explains convergence toward hybrid systems in practice.

All results are machine-checked in Lean 4 (6,000+ lines, 0 sorry).

Keywords: identification capacity, query complexity, rate-distortion, matroid structure, classification systems

I. INTRODUCTION

A. The Identification Problem

Consider an encoder-decoder pair communicating about entities from a large universe \mathcal{V} . The decoder must *identify* each entity—determine which of k classes it belongs to—using only:

- A *tag* of L bits stored with the entity, and/or
- *Queries* to a binary oracle: “does entity v satisfy attribute I ?“

This is not reconstruction (the decoder need not recover v), but *identification* in the sense of Ahlsweide and Dueck [1]: the decoder must answer “which class?” with zero or bounded error.

We prove three results:

T. Simas is with McGill University, Montreal, QC, Canada (e-mail: tristan.simas@mail.mcgill.ca).

Manuscript received January 15, 2026.

© 2026 Tristan Simas. This work is licensed under CC BY 4.0. License: <https://creativecommons.org/licenses/by/4.0/>

- 1) **Information barrier (capacity limit).** When the attribute profile $\pi : \mathcal{V} \rightarrow \{0, 1\}^n$ is not injective on classes, the channel $C \rightarrow \pi(V)$ has $I(C; \pi(V)) < H(C)$. Zero-error identification via queries alone is impossible.
- 2) **Optimal tagging (achievability).** A tag of $L = \lceil \log_2 k \rceil$ bits achieves zero-error identification with $W = O(1)$ query cost. This is Pareto-optimal in the (L, W, D) tradeoff space.
- 3) **Matroid structure (query complexity).** Minimal sufficient query sets form the bases of a matroid. The *identification dimension*—common cardinality of all minimal sets—lower-bounds the query cost W for any tag-free scheme.

These results are universal: the theory applies to type systems, databases, biological taxonomy, and knowledge graphs. We develop the mathematics in full generality, then exhibit concrete instantiations.

B. The Observation Model

We formalize the observational constraint as a family of binary predicates. The terminology is deliberately abstract; concrete instantiations follow in Section VIII.

Definition I.1 (Entity space and attribute family). Let \mathcal{V} be a set of entities (program objects, database records, biological specimens, library items). Let \mathcal{I} be a finite set of *attributes*—observable properties that partition the entity space.

Remark I.2 (Terminology). We use “attribute” for the abstract concept. In type systems, attributes are *interfaces* or *method signatures*. In databases, they are *columns*. In taxonomy, they are *phenotypic characters*. In library science, they are *facets*. The mathematics is identical.

Definition I.3 (Interface observation family). For each $I \in \mathcal{I}$, define the interface-membership observation $q_I : \mathcal{V} \rightarrow \{0, 1\}$:

$$q_I(v) = \begin{cases} 1 & \text{if } v \text{ satisfies interface } I \\ 0 & \text{otherwise} \end{cases}$$

Let $\Phi_{\mathcal{I}} = \{q_I : I \in \mathcal{I}\}$ denote the interface observation family.

Definition I.4 (Interface profile). The interface profile function $\pi : \mathcal{V} \rightarrow \{0, 1\}^{|\mathcal{I}|}$ maps each value to its complete interface signature:

$$\pi(v) = (q_I(v))_{I \in \mathcal{I}}$$

Definition I.5 (Interface indistinguishability). Values $v, w \in \mathcal{V}$ are *interface-indistinguishable*, written $v \sim w$, iff $\pi(v) = \pi(w)$.

The relation \sim is an equivalence relation. We write $[v]_\sim$ for the equivalence class of v .

Definition I.6 (Interface-only observer). An *interface-only observer* is any procedure whose interaction with a value $v \in \mathcal{V}$ is limited to queries in $\Phi_{\mathcal{I}}$. Formally, the observer receives only $\pi(v)$, not v itself.

C. The Central Question

The central question is: **what semantic properties can an interface-only observer compute?**

A semantic property is a function $P : \mathcal{V} \rightarrow \{0, 1\}$ (or more generally, $P : \mathcal{V} \rightarrow Y$ for some codomain Y). We say P is *interface-computable* if there exists a function $f : \{0, 1\}^{|\mathcal{I}|} \rightarrow Y$ such that $P(v) = f(\pi(v))$ for all v .

D. The Information Barrier

Theorem I.7 (Information barrier). *Let $P : \mathcal{V} \rightarrow Y$ be any function. If P is interface-computable, then P is constant on \sim -equivalence classes:*

$$v \sim w \implies P(v) = P(w)$$

Equivalently: no interface-only observer can compute any property that varies within an equivalence class.

Proof. Suppose P is interface-computable via f , i.e., $P(v) = f(\pi(v))$ for all v . Let $v \sim w$, so $\pi(v) = \pi(w)$. Then:

$$P(v) = f(\pi(v)) = f(\pi(w)) = P(w)$$

■

Remark I.8 (Information-theoretic nature). The barrier is *informational*, not computational. Given unlimited time, memory, and computational power, an interface-only observer still cannot distinguish v from w when $\pi(v) = \pi(w)$. The constraint is on the evidence itself.

Corollary I.9 (Provenance is not interface-computable). *Let $\text{type} : \mathcal{V} \rightarrow \mathcal{T}$ be the type assignment function. If there exist values v, w with $\pi(v) = \pi(w)$ but $\text{type}(v) \neq \text{type}(w)$, then type identity is not interface-computable.*

Proof. Direct application of Theorem I.7 to $P = \text{type}$. ■

E. The Positive Result: Nominal Tagging

We now show that augmenting interface observations with a single primitive—nominal-tag access—achieves constant witness cost.

Definition I.10 (Nominal-tag access). A *nominal tag* is a value $\tau(v) \in \mathcal{T}$ associated with each $v \in \mathcal{V}$, representing the type identity of v . The *nominal-tag access* operation returns $\tau(v)$ in $O(1)$ time.

Definition I.11 (Primitive query set). The extended primitive query set is $\Phi_{\mathcal{I}}^+ = \Phi_{\mathcal{I}} \cup \{\tau\}$, where τ denotes nominal-tag access.

Definition I.12 (Witness cost). Let $W(P)$ denote the minimum number of primitive queries from $\Phi_{\mathcal{I}}^+$ required to compute property P :

$$W(P) = \min\{c(A) : A \text{ is a procedure computing } P\}$$

where $c(A)$ counts the number of queries to $\Phi_{\mathcal{I}}^+$ made by A .

Theorem I.13 (Constant witness for type identity). *Under nominal-tag access, type identity checking has constant witness cost:*

$$W(\text{type-identity}) = O(1)$$

Specifically, the witness procedure is: return $\tau(v_1) = \tau(v_2)$.

Proof. The procedure makes exactly 2 primitive queries (one τ access per value) and one comparison. This is $O(1)$ regardless of the number of interfaces $|\mathcal{I}|$. ■

Theorem I.14 (Interface-only lower bound). *For interface-only observers, type identity checking requires:*

$$W(\text{type-identity}) = \Omega(|\mathcal{I}|)$$

in the worst case.

Proof. Construct a family of $|\mathcal{I}|$ types where each type T_i satisfies exactly the interfaces $\{I_1, \dots, I_i\}$. Distinguishing T_i from T_{i+1} requires querying I_{i+1} . Thus, distinguishing all pairs requires querying all $|\mathcal{I}|$ interfaces. ■

F. Main Contributions

This paper establishes the following results:

- 1) **Information Barrier Theorem** (Theorem I.7): Interface-only observers cannot compute any property that varies within \sim -equivalence classes. This is an information-theoretic impossibility, not a computational limitation.
- 2) **Constant-Witness Theorem** (Theorem I.13): Nominal-tag access achieves $W(\text{type-identity}) = O(1)$, with matching lower bound $\Omega(|\mathcal{I}|)$ for interface-only observers (Theorem I.14).
- 3) **Complexity Separation** (Section III): We establish $O(1)$ vs $O(k)$ vs $\Omega(n)$ complexity bounds for error localization under different observation regimes.
- 4) **Matroid Structure** (Section V): Minimal distinguishing query sets form the bases of a matroid. All such sets have equal cardinality, establishing a well-defined “distinguishing dimension.”
- 5) **(L, W, D) Optimality** (Section VII): Nominal-tag observers achieve the unique Pareto-optimal point in the (L, W, D) tradeoff space (tag length, witness cost, distortion).
- 6) **Machine-Checked Proofs:** All results formalized in Lean 4 (6,086 lines, 265 theorems, 0 sorry placeholders).

G. Related Work and Positioning

The information barrier (Theorem I.7) is related to results in query complexity and communication complexity, where limited observations constrain computable functions. The matroid structure connects to lattice-theoretic approaches in abstract interpretation [2].

The (L, W, D) analysis extends classical rate-distortion theory [3], [4] to a discrete classification setting with three dimensions: tag length L , witness cost W (query complexity), and semantic distortion D (fidelity).

Historical context. In programming language theory, the relationship between “duck typing” (attribute-only observation) and nominal typing has been discussed since Smalltalk (1980) and formalized in work on structural vs. nominal subtyping [5]. Proponents of structural typing argue that if two entities “walk like a duck and quack like a duck,” they should be treated identically. Proponents of nominal typing argue that provenance matters.

Within our observation model, we prove a quantitative separation: duck typing incurs $\Omega(n)$ witness cost where nominal tagging achieves $O(1)$. This is a machine-checked theorem (Lean 4, 0 sorry). The result does not “resolve” the debate—which involves usability, expressiveness, and tooling concerns beyond this model—but it does establish that the tradeoff has a precise information-theoretic component.

Practical implications. Programming languages have independently converged on hybrid classification: Python added Abstract Base Classes (PEP 3119), TypeScript introduced branded types, Rust’s trait system combines structural interfaces with nominal identity. This convergence is consistent with the information-theoretic analysis: nominal tags provide $O(1)$ identification at the cost of $O(\log k)$ bits of storage. The same tradeoff appears in database schema design (primary keys), biological taxonomy (species identifiers), and knowledge representation (entity URIs).

The contribution is not advocacy for a particular language feature, but a formal framework for analyzing the identification-cost tradeoff in any classification system.

H. Paper Organization

Section II formalizes the compression framework and defines the (L, W, D) tradeoff. Section III establishes complexity bounds for error localization. Section V proves the matroid structure of type axes. Section VI analyzes witness cost in detail. Section VII proves Pareto optimality. Section VIII instantiates the theory in real runtimes. Section IX concludes. Appendix A describes the Lean 4 formalization.

II. COMPRESSION FRAMEWORK

A. Semantic Compression: The Problem

The fundamental problem of *semantic compression* is: given a value v from a large space \mathcal{V} , how can we represent v compactly while preserving the ability to answer semantic queries about v ? This differs from classical source coding in that the goal is not reconstruction but *identification*—determining which equivalence class v belongs to.

Classical rate-distortion theory [3] studies the tradeoff between representation size and reconstruction fidelity. We extend this to a discrete classification setting with three dimensions: *tag length* L (bits of storage), *witness cost* W (interactive identification cost—queries or bits of communication required to determine class membership), and *distortion* D (semantic fidelity).

B. Universe of Discourse

Definition II.1 (Classification scheme). A *classification scheme* is any procedure (deterministic or randomized), with arbitrary time and memory, whose only access to a value $v \in \mathcal{V}$ is via:

- 1) The *observation family* $\Phi = \{q_I : I \in \mathcal{I}\}$, where $q_I(v) = 1$ iff v satisfies attribute I ; and optionally
- 2) A *nominal-tag primitive* $\tau : \mathcal{V} \rightarrow \mathcal{T}$ returning an opaque type identifier.

All theorems in this paper quantify over all such schemes.

This definition is intentionally broad: schemes may be adaptive, randomized, or computationally unbounded. The constraint is *observational*, not computational.

Theorem II.2 (Information barrier). *For all classification schemes with access only to Φ (no nominal tag), the output is constant on \sim_Φ -equivalence classes. Therefore, no such scheme can compute any property that varies within a \sim_Φ -class.*

Proof. Let $v \sim_\Phi w$, meaning $q_I(v) = q_I(w)$ for all $I \in \mathcal{I}$. Any scheme’s execution trace depends only on query responses. Since all queries return identical values for v and w , the scheme cannot distinguish them. Any output must therefore be identical. ■

Proposition II.3 (Model capture). *Any real-world classification protocol whose evidence consists solely of attribute-membership queries is representable as a scheme in the above model. Conversely, any additional capability corresponds to adding new observations to Φ .*

This proposition forces any objection into a precise form: to claim the theorem does not apply, one must name the additional observation capability not in Φ . “Different universe” is not a coherent objection—it must reduce to “I have access to oracle $X \notin \Phi$.”

C. The Two-Axis Model

We adopt a two-axis model of semantic structure, where each value is characterized by:

- **Bases axis (B):** The inheritance lineage—which types the value inherits from
- **Structure axis (S):** The interface signature—which methods/attributes the value provides

Definition II.4 (Two-axis representation). A value $v \in \mathcal{V}$ has representation $(B(v), S(v))$ where:

$$B(v) = \text{MRO}(\text{type}(v)) \quad (\text{Method Resolution Order}) \quad (1)$$

$$S(v) = \pi(v) = (q_I(v))_{I \in \mathcal{I}} \quad (\text{interface profile}) \quad (2)$$

Theorem II.5 (Model completeness). *In any class system with explicit inheritance, the pair (B, S) is complete: every semantic property of a value is a function of $(B(v), S(v))$.*

Proof. The proof proceeds by showing that any additional property (e.g., module location, metadata) is either derived from (B, S) or stored within the namespace (part of S). In Python, `type(name, bases, namespace)` is the universal type constructor, making (B, S) constitutive. ■

D. Interface Equivalence and Observational Limits

Recall from Section 1 the interface equivalence relation:

Definition II.6 (Interface equivalence (restated)). Values $v, w \in \mathcal{V}$ are interface-equivalent, written $v \sim w$, iff $\pi(v) = \pi(w)$ —i.e., they satisfy exactly the same interfaces.

Proposition II.7 (Equivalence class structure). *The relation \sim partitions \mathcal{V} into equivalence classes. Let \mathcal{V}/\sim denote the quotient space. An interface-only observer effectively operates on \mathcal{V}/\sim , not \mathcal{V} .*

Corollary II.8 (Information loss quantification). *The information lost by interface-only observation is:*

$$H(\mathcal{V}) - H(\mathcal{V}/\sim) = H(\mathcal{V}|\pi)$$

where H denotes entropy. This quantity is positive whenever multiple types share the same interface profile.

E. Witness Cost: Query Complexity for Semantic Properties

Definition II.9 (Witness procedure). A *witness procedure* for property $P : \mathcal{V} \rightarrow Y$ is an algorithm A that:

- 1) Takes as input a value $v \in \mathcal{V}$ (via query access only)
- 2) Makes queries to the primitive set $\Phi_{\mathcal{I}}^+$
- 3) Outputs $P(v)$

Definition II.10 (Witness cost). The *witness cost* of property P is:

$$W(P) = \min_{A \text{ computes } P} c(A)$$

where $c(A)$ is the worst-case number of primitive queries made by A .

Remark II.11 (Relationship to query complexity). Witness cost is a form of query complexity [6] specialized to semantic properties. Unlike Kolmogorov complexity, W is computable and depends on the primitive set, not a universal machine.

Lemma II.12 (Witness cost lower bounds). *For any property P :*

- 1) If P is interface-computable: $W(P) \leq |\mathcal{I}|$
- 2) If P varies within some \sim -class: $W(P) = \infty$ for interface-only observers
- 3) With nominal-tag access: $W(\text{type-identity}) = O(1)$

F. The (L, W, D) Tradeoff

We now define the three-dimensional tradeoff space that characterizes observation strategies, using information-theoretic units.

Definition II.13 (Tag length (Rate)). Let \mathcal{T} be the set of type identifiers (tags) with $|\mathcal{T}| = k$. The *tag length* L is the number of bits required to encode a type identifier:

$$L \geq \log_2 k \text{ bits per value}$$

For nominal-tag observers, $L = \lceil \log_2 k \rceil$ (optimal prefix-free encoding). For interface-only observers, $L = 0$ (no explicit tag stored). Under a distribution P over types, the expected tag length is $\mathbb{E}[L] \geq H(P)$ by Shannon's source coding theorem [3].

Definition II.14 (Witness cost (Query/Communication complexity)). The *witness cost* W is the minimum number of primitive queries (or bits of interactive communication) required for type identity checking:

$$W = \min_{A \text{ decides type-identity}} c(A)$$

where $c(A)$ is the worst-case query count. This is a form of query complexity [6] or interactive identification cost.

Definition II.15 (Behavioral equivalence). Let $\text{behavior} : \mathcal{V} \rightarrow \mathcal{B}$ be a function mapping each entity to its observable behavior—the set of responses to all possible operations. Two entities v, w are *behaviorally equivalent*, written $v \equiv w$, iff $\text{behavior}(v) = \text{behavior}(w)$.

In type systems, $\text{behavior}(v)$ is the denotational semantics: the function computed by v . In databases, it is the set of query results. In taxonomy, it is the phenotype. The formalism is parametric in this choice.

Definition II.16 (Distortion function). Let $d : \mathcal{V} \times \mathcal{V} \rightarrow \{0, 1\}$ be the misclassification indicator:

$$d(v, \hat{v}) = \begin{cases} 0 & \text{if } \text{type}(v) = \text{type}(\hat{v}) \Rightarrow v \equiv \hat{v} \\ 1 & \text{otherwise (semantic error)} \end{cases}$$

That is, $d = 0$ when type equality implies behavioral equivalence (soundness), and $d = 1$ when the observer conflates behaviorally distinct entities.

Definition II.17 (Expected and worst-case distortion). Given a distribution P over values, the *expected distortion* is:

$$D = \mathbb{E}_{v \sim P}[d(v, \hat{v})]$$

The *zero-error regime* requires $D = 0$ (no semantic errors for any v). All theorems in this paper are proved in the zero-error regime—the strongest case, where the separation is sharpest.

Remark II.18 (Distortion interpretation). $D = 0$ means the observation strategy is *sound*: type equality (as computed by the observer) implies behavioral equivalence. $D > 0$ means the strategy may conflate behaviorally distinct values with positive probability.

G. The (L, W, D) Tradeoff Space

Admissible schemes. To make the Pareto-optimality claim precise, we specify the class of admissible observation strategies:

- **Deterministic or randomized:** Schemes may use randomness; W is worst-case query count.

- **Computationally unbounded:** No time/space restrictions; the constraint is observational.
- **No preprocessing over type universe:** The scheme cannot precompute a global lookup table indexed by all possible types.
- **Tags are injective on classes:** A nominal tag $\tau(v)$ uniquely identifies the type of v . Variable-length or compressed tags are permitted; L counts bits.
- **No amortization across queries:** W is per-identification cost, not amortized over a sequence.

Justification. The “no preprocessing” and “no amortization” constraints exclude trivializations:

- *Preprocessing:* With unbounded preprocessing over the type universe \mathcal{T} , one could build a lookup table mapping attribute profiles to types. This reduces identification to $O(1)$ table lookup—but the table has size $O(|\mathcal{T}|)$, hiding the complexity in space rather than eliminating it. The constraint models systems that cannot afford $O(|\mathcal{T}|)$ storage per observer.
- *Amortization:* If W were amortized over n identifications, one could cache earlier results. This again hides complexity in state. The per-identification model captures stateless observers (typical in type checking, database queries, and biological identification).

Dropping these constraints changes the achievable region but not the qualitative separation: nominal tags still dominate for $D = 0$ because they provide $O(1)$ worst-case identification without requiring $O(|\mathcal{T}|)$ preprocessing.

Under these rules, “dominance” means strict improvement on at least one of (L, W, D) with no regression on others.

Definition II.19 (Achievable region). A point (L, W, D) is *achievable* if there exists an admissible observation strategy realizing those values. Let $\mathcal{R} \subseteq \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times [0, 1]$ denote the achievable region.

Definition II.20 (Pareto optimality). A point (L^*, W^*, D^*) is *Pareto-optimal* if there is no achievable (L, W, D) with $L \leq L^*$, $W \leq W^*$, $D \leq D^*$, and at least one strict inequality.

The main result of Section VII is that nominal-tag observation achieves the unique Pareto-optimal point with $D = 0$.

H. Converse: Tag Rate Lower Bound

Theorem II.21 (Converse). Any scheme achieving $D = 0$ (zero-error identification) requires tag length $L \geq \log_2 k$, where k is the number of distinct classes.

Proof. Zero error requires that distinct classes map to distinct decoder outputs. With k classes, at least $\log_2 k$ bits are needed to encode k distinct values. Without tags ($L = 0$), the decoder sees only query responses $\pi(v) \in \{0, 1\}^{|\mathcal{I}|}$. If π is not injective on classes (i.e., \exists classes $c_1 \neq c_2$ with identical profiles), zero error is impossible. The converse is tight: $L = \lceil \log_2 k \rceil$ suffices. ■

I. Rate-Distortion Tradeoff

When $D > 0$ is permitted, we obtain a classic rate-distortion tradeoff:

Proposition II.22 (Lossy identification). For any $\epsilon \in (0, 1)$, there exist schemes with $L = 0$ (no tags) achieving $D \leq \epsilon$ with query cost $W = O(\log(1/\epsilon) \cdot d)$, where d is the identification dimension. The tradeoff: smaller D requires larger W .

The zero-error corner ($D = 0$) is special: nominal tagging is the unique Pareto optimum. Relaxing to $D > 0$ enables tag-free schemes at the cost of increased query complexity. This mirrors the classical distinction between zero-error and ϵ -error capacity in channel coding.

J. Concrete Example

Worked example. Consider a type system with $k = 1000$ classes, each characterized by a subset of $n = 50$ interface methods.

Strategy	Tag L	Witness W
Nominal (class ID)	$\lceil \log_2 1000 \rceil = 10$ bits	$O(1)$
Duck typing (query all)	0	≤ 50 queries
Adaptive duck typing	0	$\geq d$ queries

Here d is the identification dimension—the size of any minimal distinguishing query set. For typical type hierarchies, $d \approx 5\text{--}15$. The gap between 10 bits of storage vs. 5–50 queries per identification is the cost of forgoing nominal tagging. At scale (millions of identifications), the cumulative query cost dominates.

III. COMPLEXITY BOUNDS

IV. CORE THEOREMS

A. The Error Localization Theorem

Definition IV.1 (Error location count). Let $E(\mathcal{O})$ be the number of locations that must be inspected to find all potential violations of a constraint under observation family \mathcal{O} .

Theorem IV.2 (Nominal-tag localization). $E(\text{nominal-tag}) = O(1)$.

Proof. Under nominal-tag observation, the constraint “ v must be of class A ” is satisfied iff $\tau(v) \in \text{subtypes}(A)$. This is determined at a single location: the definition of $\tau(v)$ ’s class. One location. ■

Theorem IV.3 (Declared interface localization). $E(\text{interface-only, declared}) = O(k)$ where $k = \text{number of entity classes}$.

Proof. With declared interfaces, the constraint “ v must satisfy interface I ” requires verifying that each class implements all attributes in I . For k classes, $O(k)$ locations. ■

Theorem IV.4 (Attribute-only localization). $E(\text{attribute-only}) = \Omega(n)$ where $n = \text{number of query sites}$.

Proof. Under attribute-only observation, each query site independently checks “does v have attribute a ?” with no centralized declaration. For n query sites, each must be inspected. Lower bound is $\Omega(n)$. ■

Corollary IV.5 (Strict dominance). *Nominal-tag observation strictly dominates attribute-only: $E(\text{nominal-tag}) = O(1) < \Omega(n) = E(\text{attribute-only})$ for all $n > 1$.*

B. The Information Scattering Theorem

Definition IV.6 (Constraint encoding locations). Let $I(\mathcal{O}, c)$ be the set of locations where constraint c is encoded under observation family \mathcal{O} .

Theorem IV.7 (Attribute-only scattering). *For attribute-only observation, $|I(\text{attribute-only}, c)| = O(n)$ where $n = \text{query sites using constraint } c$.*

Proof. Each attribute query independently encodes the constraint. No shared reference exists. Constraint encodings scale with query sites. ■

Theorem IV.8 (Nominal-tag centralization). *For nominal-tag observation, $|I(\text{nominal-tag}, c)| = O(1)$.*

Proof. The constraint “must be of class A ” is encoded once in the definition of A . All tag checks reference this single definition. ■

Corollary IV.9 (Maintenance entropy). *Attribute-only observation maximizes maintenance entropy; nominal-tag observation minimizes it.*

V. MATROID STRUCTURE

A. Query Families and Distinguishing Sets

The classification problem is: given a set of queries, which subsets suffice to distinguish all entities?

Definition V.1 (Query family). Let \mathcal{Q} be the set of all primitive queries available to an observer. For a classification system with interface set \mathcal{I} , we have $\mathcal{Q} = \{q_I : I \in \mathcal{I}\}$ where $q_I(v) = 1$ iff v satisfies interface I .

Definition V.2 (Distinguishing set). A subset $S \subseteq \mathcal{Q}$ is *distinguishing* if, for all values v, w with $\text{type}(v) \neq \text{type}(w)$, there exists $q \in S$ such that $q(v) \neq q(w)$.

Definition V.3 (Minimal distinguishing set). A distinguishing set S is *minimal* if no proper subset of S is distinguishing.

B. Matroid Structure of Query Families

Structural assumptions. The matroid theorem below is *unconditional* given the definitions above. It depends only on:

- 1) Queries are binary-valued functions $q : \mathcal{V} \rightarrow \{0, 1\}$.
- 2) “Distinguishing” is defined by: $\exists q \in S$ such that $q(v) \neq q(w)$.
- 3) “Minimal” means no proper subset suffices.

No further assumptions on the query family, value space, or type structure are required. The proof constructs a closure operator satisfying extensivity, monotonicity, and idempotence, from which basis exchange follows (see Lean formalization).

Definition V.4 (Bases family). Let $E = \mathcal{Q}$ be the ground set of all queries. Let $\mathcal{B} \subseteq 2^E$ be the family of minimal distinguishing sets.

Lemma V.5 (Basis exchange). *For any $B_1, B_2 \in \mathcal{B}$ and any $q \in B_1 \setminus B_2$, there exists $q' \in B_2 \setminus B_1$ such that $(B_1 \setminus \{q\}) \cup \{q'\} \in \mathcal{B}$.*

Proof sketch. Define the closure operator $\text{cl}(X) = \{q : X\text{-equivalence implies } q\text{-equivalence}\}$. This satisfies:

- 1) *Extensive:* $X \subseteq \text{cl}(X)$ (if $q \in X$, then X -equivalence trivially implies q -equivalence).
- 2) *Monotone:* $X \subseteq Y \Rightarrow \text{cl}(X) \subseteq \text{cl}(Y)$ (more queries give finer equivalence).
- 3) *Idempotent:* $\text{cl}(\text{cl}(X)) = \text{cl}(X)$ (closure is transitive).

A closure operator satisfying these axioms plus the *exchange property*—if $q \notin \text{cl}(X)$ and $q \in \text{cl}(X \cup \{q'\})$, then $q' \in \text{cl}(X \cup \{q\})$ —defines a matroid. The exchange property follows from symmetry of indistinguishability: if adding q' to X newly determines q , there exist witnesses v, w with X -equivalence, q' -inequivalence, and q -inequivalence; the same witnesses show adding q newly determines q' .

Minimal distinguishing sets are exactly the bases of this matroid. Full machine-checked proof: `proofs/abstract_class_system.lean`, namespace `AxisClosure`. ■

Theorem V.6 (Matroid bases). *\mathcal{B} is the set of bases of a matroid on ground set E .*

Proof. By the basis-exchange lemma and the standard characterization of matroid bases [7]. ■

Corollary V.7 (Well-defined distinguishing dimension). *All minimal distinguishing sets have equal cardinality. We call this the distinguishing dimension of the classification system.*

C. Implications for Witness Cost

Corollary V.8 (Lower bound on interface-only witness cost). *For any interface-only observer, $W(\text{type-identity}) \geq d$ where d is the distinguishing dimension.*

Proof. Any witness procedure must query at least one minimal distinguishing set. ■

The key insight: the distinguishing dimension is invariant across all minimal query strategies. The difference between nominal-tag and interface-only observers lies in *witness cost*: a nominal tag achieves $W = O(1)$ by storing the identity directly, bypassing query enumeration.

VI. WITNESS COST ANALYSIS

A. Witness Cost for Type Identity

Recall from Section 2 that the witness cost $W(P)$ is the minimum number of primitive queries required to compute property P . For type identity, we ask: what is the minimum number of queries to determine if two values have the same type?

Theorem VI.1 (Nominal-Tag Observers Achieve Minimum Witness Cost). *Nominal-tag observers achieve the minimum witness cost for type identity:*

$$W(\text{type identity}) = O(1)$$

Specifically, the witness is a single tag read: compare $\text{tag}(v_1) = \text{tag}(v_2)$.

Interface-only observers require $W(\text{type identity}) = \Omega(n)$ where n is the number of interfaces.

Proof. See Lean formalization: proofs/nominal_resolution.lean. The proof shows:

- 1) Nominal-tag access is a single primitive query
- 2) Interface-only observers must query n interfaces to distinguish all types
- 3) No shorter witness exists for interface-only observers (by the information barrier)

■

B. Witness Cost Comparison

Observer Class	Witness Procedure	Witness Cost W
Nominal-tag	Single tag read	$O(1)$
Interface-only	Query n interfaces	$O(n)$

TABLE I

WITNESS COST FOR TYPE IDENTITY BY OBSERVER CLASS.

The Lean 4 formalization (Appendix A) provides a machine-checked proof that nominal-tag access minimizes witness cost for type identity.

VII. (L, W, D) OPTIMALITY

A. Three-Dimensional Tradeoff: Tag Length, Witness Cost, Distortion

Recall from Section 2 that observer strategies are characterized by three dimensions:

- **Tag length L :** bits required to encode a type identifier ($L \geq \log_2 k$ for k types)
- **Witness cost W :** minimum number of primitive queries for type identity checking
- **Distortion D :** expected misclassification rate ($D = 0$ in the zero-error regime)

We compare two observer classes:

Definition VII.1 (Interface-only observer). An observer that queries only interface membership ($q_I \in \Phi_I$), with no access to explicit type tags.

Definition VII.2 (Nominal-tag observer). An observer that may read a single type identifier (nominal tag) per value, in addition to interface queries.

Theorem VII.3 (Pareto Optimality of Nominal-Tag Observers). *Nominal-tag observers achieve the unique Pareto-optimal point in the (L, W, D) space with $D = 0$:*

- **Tag length:** $L = \lceil \log_2 k \rceil$ bits for k types
- **Witness cost:** $W = O(1)$ queries (one tag read)
- **Distortion:** $D = 0$ (type equality implies behavior equivalence)

Interface-only observers achieve:

- **Tag length:** $L = 0$ bits (no explicit tag)
- **Witness cost:** $W = \Omega(n)$ queries (must query n interfaces)

- **Distortion:** $D > 0$ (may conflate behaviorally distinct types)

Proof. See Lean formalization: proofs/python_instantiation.lean. The proof verifies:

- 1) nominal_cost_constant: Nominal-tag achieves $(L, W, D) = (O(1), O(1), 0)$
- 2) interface_cost_linear: Interface-only requires $O(n)$ queries
- 3) python_gap_unbounded: The cost gap is unbounded in the limit
- 4) Interface observations alone cannot distinguish provenance; nominal tags can

B. Pareto Frontier

The three-dimensional frontier shows:

- Nominal-tag observers dominate interface-only observers on all three dimensions
- Interface-only observers trade tag length for distortion (zero L , but $D = 1$)

The Lean 4 formalization (Appendix A) provides a machine-checked proof of Pareto optimality for nominal-tag observers in the (L, W, D) tradeoff.

Remark VII.4 (Programming language instantiations). In programming language terms: *nominal typing* corresponds to nominal-tag observers (e.g., CPython’s `isinstance`, Java’s `.getClass()`). *Duck typing* corresponds to interface-only observers (e.g., Python’s `hasattr`). *Structural typing* is an intermediate case with $D = 0$ but $W = O(n)$.

VIII. INSTANTIATIONS IN REAL RUNTIMES

The preceding sections established abstract results about observer classes and witness cost. We now ground these in concrete systems across multiple domains, showing that real classification systems instantiate the theoretical categories—and that the complexity bounds are not artifacts of the model but observable properties of deployed implementations.

A. Biological Taxonomy: Phenotype vs Genotype

Linnean taxonomy classifies organisms by observable phenotypic characters: morphology, behavior, habitat. This is attribute-only observation. The information barrier applies: phenotypically identical organisms from distinct species are indistinguishable.

The cryptic species problem: Cryptic species share identical phenotypic profiles but are reproductively isolated and genetically distinct. Attribute-only observation (morphology) cannot distinguish them— $\pi(A) = \pi(B)$ but $\text{species}(A) \neq \text{species}(B)$.

The nominal tag: DNA barcoding provides the resolution [8]. A short genetic sequence (e.g., mitochondrial COI) acts as the nominal tag: $O(1)$ identity verification via sequence comparison. This reduced cryptic species identification from $\Omega(n)$ morphological examination to constant-time molecular lookup.

B. Library Classification: Subject vs ISBN

Library classification systems like Dewey Decimal observe subject matter—a form of attribute-only classification. Two books on the same subject are indistinguishable by subject code alone.

The nominal tag: The ISBN (International Standard Book Number) is the nominal tag [9]. Given two physical books, identity verification is $O(1)$: compare ISBNs. Without ISBNs, distinguishing two copies of different editions on the same subject requires $O(n)$ attribute inspection (publication date, page count, publisher, etc.).

C. Database Systems: Columns vs Primary Keys

Relational databases observe entities via column values. The information barrier applies: rows with identical column values (excluding the key) are indistinguishable.

The nominal tag: The primary key is the nominal tag [10]. Entity identity is $O(1)$: compare keys. This is why database theory requires keys—without them, the system cannot answer “is this the same entity?”

Natural vs surrogate keys: Natural keys (composed of attributes) are attribute-only observation and inherit its limitations. Surrogate keys (auto-increment IDs, UUIDs) are pure nominal tags: no semantic content, pure identity.

D. Programming Language Runtimes

Type systems are the motivating example for this work. We survey four runtimes.

1) **CPython: The `ob_type` Pointer:** Every CPython heap object begins with a `PyObject` header containing an `ob_type` pointer to its type object [11]. This is the nominal tag: a single machine word encoding complete type identity.

Witness procedure: Given objects `a` and `b`, type identity is `type(a) is type(b)`—two pointer dereferences and one pointer comparison. Cost: $O(1)$ primitive operations, independent of interface count.

Contrast with `hasattr`: Interface-only observation in Python uses `hasattr(obj, name)` for each required method. To verify an object satisfies a protocol with k methods requires k attribute lookups. Worse: different call sites may check different subsets, creating $\Omega(n)$ total checks where n is the number of call sites. The nominal tag eliminates this entirely.

2) **Java: `.getClass()` and the Method Table:** Java’s object model stores a pointer to the class object in every instance header [12]. The `.getClass()` method exposes this [13], and `instanceof` checks traverse the class hierarchy.

Key observation: Java’s `instanceof` is $O(d)$ where d is inheritance depth, not $O(|\mathcal{I}|)$ where $|\mathcal{I}|$ is the number of interfaces. This is because `instanceof` walks the inheritance hierarchy (a nominal-tag query), not the interface list (an attribute query).

3) **TypeScript: Structural Equivalence:** TypeScript uses attribute-only (declared) observation [14]: the compiler checks structural compatibility, not nominal identity. Two types are assignment-compatible iff their structures match.

Implication: Type identity checking requires traversing the structure. For a type with n fields/methods, $W(\text{type-identity}) = O(n)$. This is inherent to the observation model: no compilation strategy can reduce this to $O(1)$ without adding nominal tags.

4) **Rust: Static Nominal Tags:** Rust resolves type identity at compile time via its nominal type system. At runtime, `std::any::TypeId` provides nominal-tag access [15].

The dyn Trait case: Rust’s trait objects include a vtable pointer but not a type tag [16]. This is attribute-only observation: the vtable encodes which methods exist, not which type provided them.

E. Cross-Domain Summary

Domain	Attribute-Only	Nominal Tag	W
Biology	Phenotype (morphology)	DNA barcode (COI)	$O(1)$
Libraries	Subject (Dewey)	ISBN	$O(1)$
Databases	Column values	Primary key	$O(1)$
CPython	hasattr probing	<code>ob_type</code> pointer	$O(1)$
Java	Interface check	<code>.getClass()</code>	$O(1)$
TypeScript	Structural check	(none at runtime)	$O(n)$
Rust (static)	Trait bounds	<code>TypeId</code>	$O(1)$

TABLE II
WITNESS COST FOR IDENTITY ACROSS CLASSIFICATION SYSTEMS.
NOMINAL TAGS ACHIEVE $O(1)$; ATTRIBUTE-ONLY PAYS $O(n)$ OR $O(k)$.

The pattern is universal: systems with nominal tags achieve $O(1)$ witness cost; systems without them pay $O(n)$ or $O(k)$. This is not domain-specific—it is the information barrier theorem instantiated across classification systems.

IX. CONCLUSION

This paper presents an information-theoretic analysis of classification under observational constraints. We prove three main results:

- 1) **Information Barrier:** Observers limited to attribute-membership queries cannot compute properties that vary within indistinguishability classes. This is universal: it applies to biological taxonomy, database systems, library classification, and programming language runtimes alike.
- 2) **Witness Optimality:** Nominal-tag observers achieve $W(\text{identity}) = O(1)$, the minimum witness cost. The gap from attribute-only observation ($\Omega(n)$) is unbounded.
- 3) **Matroid Structure:** Minimal distinguishing query sets form the bases of a matroid. The distinguishing dimension of a classification problem is well-defined and computable.

A. The Universal Pattern

Across domains, the same structure recurs:

- **Biology:** Phenotypic observation cannot distinguish cryptic species. DNA barcoding (nominal tag) resolves them in $O(1)$.
- **Databases:** Column-value queries cannot distinguish rows with identical attributes. Primary keys (nominal tag) provide $O(1)$ identity.

- **Type systems:** Interface observation cannot distinguish structurally identical types. Type tags provide $O(1)$ identity.

The information barrier is not a quirk of any particular domain—it is a mathematical necessity arising from the quotient structure induced by limited observations.

B. Implications

- **Nominal tags are not optional** when identity queries are required. They are the unique mechanism achieving $O(1)$ witness cost with zero distortion.
- **The barrier is informational, not computational:** even with unbounded resources, attribute-only observers cannot overcome it.
- **Classification system design is constrained:** the choice of observation family determines which properties are computable.

C. Future Work

- 1) **Other classification domains:** What is the matroid structure of observation spaces in chemistry (molecular fingerprints), linguistics (phonetic features), or machine learning (feature embeddings)?
- 2) **Witness complexity of other properties:** Beyond identity, what are the witness costs for provenance, equivalence, or subsumption?
- 3) **Hybrid observers:** Can observer strategies that combine tags and attributes achieve better (L, W, D) tradeoffs for specific query distributions?

D. Conclusion

Classification under observational constraints admits a clean information-theoretic analysis. Nominal tags are not a design preference—they are the provably optimal strategy for identity verification under the (L, W, D) tradeoff. The results are universal, and all proofs are machine-verified in Lean 4.

AI Disclosure

This work was developed with AI assistance (Claude, Anthropic). The AI contributed to exposition, code generation, and proof exploration. All mathematical claims were verified by the authors and machine-checked in Lean 4. The Lean proofs are the authoritative source; no theorem depends solely on AI-generated reasoning.

REFERENCES

- [1] R. Ahlswede and G. Dueck, “Identification via channels,” *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 15–29, 1989.
- [2] P. Cousot and R. Cousot, “Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints,” in *Proceedings of the 4th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages*, 1977, pp. 238–252.
- [3] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” *IRE National Convention Record*, vol. 7, pp. 142–163, 1959.
- [4] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, 1971.
- [5] L. Cardelli and P. Wegner, “On understanding types, data abstraction, and polymorphism,” *ACM Computing Surveys*, vol. 17, no. 4, pp. 471–523, 1985.

- [6] H. Buhrman, L. Fortnow, M. Koucký, and B. Loff, “Computational complexity of discrete problems,” *Bulletin of the EATCS*, vol. 77, pp. 42–56, 2002.
- [7] D. J. A. Welsh, *Matroid Theory*. Academic Press, 1976.
- [8] P. D. N. Hebert, A. Cywinski, S. L. Ball, and J. R. deWaard, “Biological identifications through dna barcodes,” pp. 313–321, 2003.
- [9] International ISBN Agency, *ISBN Users’ Manual*, 7th ed. International ISBN Agency, 2017.
- [10] E. F. Codd, *The Relational Model for Database Management: Version 2*. Addison-Wesley, 1990.
- [11] P. S. Foundation, “Cython object implementation: Object structure,” <https://github.com/python/cpython/blob/main/Include/object.h>, 2024.
- [12] Oracle, “The java virtual machine specification: Chapter 2.5 runtime data areas,” <https://docs.oracle.com/javase/specs/jvms/se21/html/jvms-2.html>, 2024.
- [13] ———, “Java object: getClass()”, <https://docs.oracle.com/javase/8/docs/api/java/lang/Object.html#getClass-->, 2024.
- [14] Microsoft, “TypeScript handbook: Type compatibility,” <https://www.typescriptlang.org/docs/handbook/type-compatibility.html>, 2024.
- [15] R. Foundation, “Rust std::any::type_id,” https://doc.rust-lang.org/std/any/fn.type_id.html, 2024.
- [16] ———, “The rust reference: Trait objects,” <https://doc.rust-lang.org/reference/types/trait-object.html>, 2024.
- [17] W. R. Cook, W. Hill, and P. S. Canning, “Inheritance is not subtyping,” in *Proceedings of the 17th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, 1990, pp. 125–135.
- [18] M. Abadi and L. Cardelli, “A theory of objects.” Springer-Verlag, 1996.
- [19] B. Liskov and J. M. Wing, “A behavioral notion of subtyping,” vol. 16, no. 6, 1994, pp. 1811–1841.
- [20] J. C. Reynolds, “Types, abstraction and parametric polymorphism,” *Information Processing*, vol. 83, pp. 513–523, 1983.

APPENDIX

We provide machine-checked proofs of our core theorems in Lean 4. The complete development (6300+ lines across eleven modules, 0 `sorry` placeholders) is organized as follows:

1. **Language-agnostic layer** (Section 6.12): The two-axis model (B, S) , axis lattice metatheorem, and strict dominance: proving nominal-tag observation dominates interface-only observation in `any` class system with explicit inheritance. These proofs require no Python-specific axioms.
2. **Python instantiation layer** (Sections 6.1–6.11): The dual-axis resolution algorithm, provenance preservation, and OpenHCS-specific invariants: proving that Python’s `type(name, bases, namespace)` and C3 linearization correctly instantiate the abstract model.
3. **Complexity bounds layer** (Section 6.13): Formalization of $O(1)$ vs $O(k)$ vs $\Omega(n)$ complexity separation. Proves that nominal error localization is $O(1)$, interface-only (declared) is $O(k)$, interface-only is $\Omega(n)$, and the gap grows without bound.

The abstract layer establishes that our theorems apply to Java, C#, Ruby, Scala, and any language with the (B, S) structure. The Python layer demonstrates concrete realization. The complexity layer proves the asymptotic dominance is machine-checkable, not informal.

A. Type Universe and Registry

Types are represented as natural numbers, capturing nominal identity:

-- Types are represented as natural numbers (nominal)
abbrev Typ := Nat

TABLE III
LEAN 4 FORMALIZATION MODULES

Module	Lines	Theorems	Purpose
abstract_class_system.lean	3082	90+	Core: two-axis model, dominance
axis_framework.lean	1667	40+	Matroid structure
nominal_resolution.lean	556	21	Resolution, capabilities
discipline_migration.lean	142	11	Discipline vs migration
context_formalization.lean	215	7	Greenfield/retrofit
python_instantiation.lean	247	12	Python instantiation
typescript_instantiation.lean	65	3	TypeScript instantiation
java_instantiation.lean	63	3	Java instantiation
rust_instantiation.lean	64	3	Rust instantiation
Total	6100+	190+	

```
-- The lazy-to-base registry as a partial function ResolveResult where
def Registry := Typ -> Option Typ
                           value : FieldValue
                           scope : ScopeId
-- A registry is well-formed if base types are not yet defined
def Registry.wellFormed (R : Registry) : Prop by DecidableEq
  forall L B, R L = some B -> R B = none
                                         -- Find first matching config in a list
-- Normalization: map lazy type to base, or def findConfigByType (configs : List ConfigInstance)
def normalizeType (R : Registry) (T : Typ) : TypOption FieldValue :=
  match R T with
  | some B => B
  | none => T
                                         match configs.find? (fun c => c.typ == T) with
                                         | some c => some c.fieldValue
                                         | none => none

Invariant (Normalization Idempotence). For well-formed registries, normalization is idempotent:
theorem normalizeType_idempotent (R : Registry) (h_wf : R.wellFormed) :
  normalizeType R (normalizeType R T) = normalizeType R T
simp only [normalizeType]
cases hR : R T with
| none => simp only [hR]
| some B =>
  have h_base : R B = none := h_wf T B hR
  simp only [h_base]

B. MRO and Scope Stack
-- MRO is a list of types, most specific first
abbrev MRO := List Typ

-- Scope stack: most specific first
abbrev ScopeStack := List ScopeId

-- Config instance: type and field value
structure ConfigInstance where
  typ : Typ
  fieldValue : FieldValue

-- Configs available at each scope
def ConfigContext := ScopeId -> List ConfigInstance

C. The RESOLVE Algorithm
-- Resolution result: value, scope, source type | none => 0

D. GETATTRIBUTE Implementation
-- Raw field access (before resolution)
def rawFieldValue (obj : ConfigInstance) : FieldValue :=
  obj.fieldValue

-- GETATTRIBUTE implementation
def getattribute (R : Registry) (obj : ConfigInstance) (scopes : ScopeStack) (ctx : ConfigContext) (i : FieldValue :=
  let raw := rawFieldValue obj
  if raw != 0 then raw -- Concrete value, no resolution needed
  else if isLazyField then
    match resolve R mro scopes ctx with
    | some result => result.value
  | none => 0

```

```
else raw
```

E. Theorem 6.1: Resolution Completeness

Theorem 6.1 (Completeness). The resolve function is complete: it returns value v if and only if either no resolution occurred ($v = 0$) or a valid resolution result exists.

```
theorem resolution_completeness
  (R : Registry) (mro : MRO)
  (scopes : ScopeStack) (ctx : ConfigContext)
  (match resolve R mro scopes ctx with
  | some r => r.value
  | none => 0) = v <->
  (v = 0 /\ $\backslash$  resolve R mro scopes ctx = none) /\ $\backslash$ 
  (exists r : ResolveResult,
    resolve R mro scopes ctx = some r /\ $\backslash$  r.value)
cases hr : resolve R mro scopes ctx with
| none =>
  constructor
  . intro h; left; exact (h.symm, rfl)
  . intro h
    rcases h with (hv, _) | (r, hfalse, _)
    . exact hv.symm
    . cases hfalse
| some result =>
  constructor
  . intro h; right; exact (result, rfl, h)
  . intro h
    rcases h with (_, hfalse) | (r, hr2, hv)
    . cases hfalse
    . simp only [Option.some.injEq] at hr2
      rw [← hr2] at hv; exact hv
```

F. Theorem 6.2: Provenance Preservation

Theorem 6.2a (Uniqueness). Resolution is deterministic: same inputs always produce the same result.

```
theorem provenance_uniqueness
  (R : Registry) (mro : MRO) (scopes : ScopeStack) (ctx : ConfigContext)
  (result_1 result_2 : ResolveResult)
  (hr_1 : resolve R mro scopes ctx = some result_1)
  (hr_2 : resolve R mro scopes ctx = some result_2)
  result_1 = result_2 := by
  simp only [hr_1, Option.some.injEq] at hr_2
  exact hr_2
```

Theorem 6.2b (Determinism). Resolution function is deterministic.

```
theorem resolution_determinism
  (R : Registry) (mro : MRO) (scopes : ScopeStack) (ctx : ConfigContext)
  forall r_1 r_2, resolve R mro scopes ctx = some r_1 /\ $\backslash$ 
  resolve R mro scopes ctx = r_2 -> r_1 = r_2 := by
  intros r_1 r_2 h_1 h_2
  rw [← h_1, ← h_2]
```

G. Duck Typing Formalization

We now formalize interface-only observation and prove it cannot provide provenance.

Duck object structure:

```
-- In interface-only observation, a "type" is just
-- There's no nominal identity - only structure matches
structure InterfaceValue where
  fields : List (String * Nat)
deriving DecidableEq
```

Structural equivalence:

```
-- Two interface-only values are "interface-equivalent"
-- This is THE defining property of interface-only values
def interfaceEquivalent (a b : InterfaceValue) : Prop
  forall name, getField a name = getField b name
```

We prove this is an equivalence relation:

```
theorem structEq_refl (a : InterfaceValue) :
  interfaceEquivalent a a := by
  intro name; rfl
```

```
theorem structEq_symm (a b : InterfaceValue) :
  interfaceEquivalent a b -> interfaceEquivalent b a
  intro h name; exact (h name).symm
```

```
theorem structEq_trans (a b c : InterfaceValue) :
  interfaceEquivalent a b -> interfaceEquivalent b c
  interfaceEquivalent a c := by
  intro hab hbc name; rw [hab name, hbc name]
```

The Interface-Equivalence Principle:

Any function operating on interface-only values must respect interface equivalence. If two objects have the same interface profile, they are indistinguishable. This follows from the definition of interface-only observation: "If it satisfies the interface, it IS considered equivalent."

```
interfaceRespecting function treats interface
  as equal if interfaceRespecting (f : InterfaceValue -> a)
  forall a b, interfaceEquivalent a b -> f a = f b
```

H. Corollary 6.3: Duck Typing Cannot Provide Provenance

Provenance requires returning WHICH object provided a value. But under interface-only observation typing, interface-equivalent objects are indistinguishable. Therefore, any "provenance" must be constant on equivalent objects.

```
Suppose we try to build a provenance function for
  It would have to return which InterfaceValue provides
  structure DuckProvenance where
    value : Nat
    source : InterfaceValue -- "Which object provided"
```

```
deriving DecidableEq
```

Theorem (Indistinguishability). Any interface-respecting provenance function cannot distinguish sources:

```
theorem interface_provenance_indistinguishable
  (getProvenance : InterfaceValue -> Option DuckProvenance)
  (h_interface : InterfaceRespecting getProvenance)
  (obj1 obj2 : InterfaceValue)
  (h_equiv : interfaceEquivalent obj1 obj2)
  getProvenance obj1 = getProvenance obj2
  exact h_interface obj1 obj2 h_equiv
```

Corollary 6.3 (Absurdity). If two objects are interface-equivalent and both provide provenance, the provenance must claim the SAME source for both (absurd if they're different objects):

```
theorem interface_provenance_absurdity
  (getProvenance : InterfaceValue -> Option DuckProvenance)
  (h_interface : InterfaceRespecting getProvenance)
  (obj1 obj2 : InterfaceValue)
  (h_equiv : interfaceEquivalent obj1 obj2)
  (prov1 prov2 : DuckProvenance)
  (h1 : getProvenance obj1 = some prov1)
  (h2 : getProvenance obj2 = some prov2) :
  prov1 = prov2 := by
  have h_eq := h_interface obj1 obj2 h_equiv
  rw [h1, h2] at h_eq
  exact Option.some.inj h_eq
```

The key insight: In interface-only observation, if `obj1` and `obj2` have the same fields, they are interface-equivalent. Any interface-respecting function returns the same result for both. Therefore, provenance CANNOT distinguish them. Therefore, provenance is IMPOSSIBLE in interface-only observation.

Contrast with nominal-tag observation: In our nominal system, types are distinguished by identity:

```
-- Example: Two nominally different types
def WellFilterConfigType : Nat := 1
def StepWellFilterConfigType : Nat := 2
```

```
-- These are distinguishable despite potentially identical structure
theorem nominal_types_distinguishable :
```

```
  WellFilterConfigType != StepWellFilterConfigType := by decide
```

Therefore, `ResolveResult.sourceType` is meaningful: it tells you WHICH type provided the value, even if types have the same structure.

I. Verification Status

All proofs compile without warnings or sorry placeholders. The verification covers:

- **AbstractClassSystem** (475 lines): Two-axis model, strict dominance, axis lattice metatheorem
- **NominalResolution** (157 lines): Resolution algorithm, completeness (Thm. 6.1), uniqueness (Thm. 6.2)
- **DuckTyping** (127 lines): Structural equivalence, impossibility (Cor. 6.3)

- **MetaprogrammingGap** (156 lines): Declaration/query model, hook requirements
- **CapabilityExhaustiveness** (42 lines): Capability completeness theorems
- **AdapterAmortization** (60 lines): Cost model, amortization proofs

Total: 6,100+ lines, 190+ theorems, 0 sorry, 0 warnings.

J. What the Lean Proofs Guarantee

`:= by`

The machine-checked verification establishes:

1. **Algorithm correctness:** `resolve` returns value `v` iff resolution found a config providing `v` (Theorem 6.1).
2. **Determinism:** Same inputs always produce same (`value, scope, sourceType`) tuple (Theorem 6.2).
3. **Idempotence:** Normalizing an already-normalized type is a no-op (`normalization_idempotent`).

4. **Interface-only observation impossibility:** Any function respecting interface equivalence cannot distinguish between interface-identical objects, making provenance tracking impossible (Corollary 6.3).

K. What the proofs do NOT guarantee:

- **C3 correctness:** We assume MRO is well-formed. Python's C3 algorithm can fail on pathological diamonds (raising `TypeError`). Our proofs apply only when C3 succeeds.
- **Registry invariants:** `Registry.wellFormed` is an axiom (base types not in domain). We prove theorems given this axiom but do not derive it from more primitive foundations.
- **Termination and complexity:** We use Lean's termination checker to verify `resolve` terminates. The complexity bound $O(|\text{scopes}| \times |\text{MRO}|)$ is also mechanically verified via `resolution_complexity_bound` and related lemmas proving linearity in each dimension.

This is standard practice in mechanized verification: ComCert assumes well-typed input, seL4 assumes hardware correctness. Our proofs establish that given a well-formed registry and MRO, the resolution algorithm is correct and provides provenance, meaning interface-only observation cannot.

K. On the Nature of Foundational Proofs

A reader examining the Lean source code will notice that most proofs are remarkably short, often 1-3 lines. For example, the provenance impossibility theorem (Theorem 3.13) has a one-line proof: `exact h_shape A B h_same_ns`. This brevity is not an accident or a sign of triviality. It is the hallmark of *foundational* work, where the insight lies in the formalization, not the derivation.

Definitional vs. derivational proofs. Our core theorems establish *definitional* impossibilities, not algorithmic complexities. When we prove that no shape-respecting function can compute provenance (Theorem 3.13), we are not saying “all known algorithms fail” or “the problem is NP-hard.” We are saying something stronger: *it is information-theoretically impossible*. The proof follows immediately from the definition of

shape-respecting functions. If two types have the same shape, any shape-respecting function must treat them identically. This is not a complex derivation; it is an unfolding of definitions.

Precedent in foundational CS. This pattern appears throughout foundational computer science:

- **Turing’s Halting Problem (1936):** The proof is a simple diagonal argument, perhaps 10 lines in modern notation. Yet it establishes a fundamental limit on computation that no future algorithm can overcome.
- **Brewer’s CAP Theorem (2000):** The impossibility proof is straightforward: if a partition occurs, a system cannot be both consistent and available. The insight is in the *formalization* of what consistency, availability, and partition-tolerance mean, not in the proof steps.
- **Curry-Howard Correspondence (1958/1969):** The isomorphism between types and propositions is almost definitional once the right abstractions are identified. The profundity is in recognizing the correspondence, not deriving it.

Why simplicity indicates strength. A definitional impossibility is *stronger* than a computational lower bound. Proving that sorting requires $\Omega(n \log n)$ comparisons in the worst case (decision tree argument) leaves open the possibility of non-comparison-based algorithms (radix sort, counting sort). Proving that provenance is not shape-respecting *closes all loopholes*. No algorithm, no external state, no future language feature can make interface-only observation compute provenance without abandoning the definition of “shape-based.”

Where the insight lies. The semantic contribution of our formalization is threefold:

- 1) **Precision forcing.** Formalizing “interface-only observation” in Lean requires stating exactly what it means for a function to be shape-respecting (Definition: ShapeRespecting). This precision eliminates ambiguity. Informal arguments can wave hands; formal proofs cannot.
- 2) **Completeness guarantee.** The query space partition (Theorem 3.19) proves that *every* query is either shape-respecting or Bases-dependent. The partition is mathematical (*tertium non datur*), deriving the capability gap from logic.
- 3) **Universal scope.** The proofs apply to *any* interface-only observation discipline, not just specific implementations. The impossibility holds for interface-only observation (Python), interface-only (declared) observation (TypeScript), Protocols (PEP 544), and any future system that discards the Bases axis.

What machine-checking guarantees. The Lean compiler verifies that every proof step is valid, every definition is consistent, and no axioms are added beyond Lean’s foundations. The proofs use Lean 4 with Mathlib and classical reasoning (open Classical), relying on Lean’s standard axioms (propext, Quot.sound, Classical.choice)—no custom axioms are introduced. Zero sorry placeholders means zero unproven claims. The 6000+ lines establish a verified chain from axioms to theorems. Reviewers need not trust our informal

explanations. They can run `lake build` and verify the proofs themselves.

Comparison to informal arguments. Prior work on typing disciplines (Cook et al. [17], Abadi & Cardelli [18]) presents compelling informal arguments but lacks machine-checked proofs. Our contribution is not new *wisdom*. The insight that nominal-tag observation provides capabilities interface-only (declared) observation lacks is old. Our contribution is *formalization*: making the argument precise enough to mechanize, closing loopholes, and proving the claims hold universally within scope.

This is the tradition of metatheory established by Liskov & Wing [19] for behavioral subtyping and Reynolds [20] for parametricity. The goal is not to prove that specific programs are correct, but to establish what is *possible* within a formal framework. Simple proofs from precise definitions are the gold standard of this work.

L. External Provenance Map Rebuttal

Objection: “Interface-only observation could provide provenance via an external map: `provenance_map : Dict[id(obj), SourceType]`.”

Rebuttal: This objection conflates *object identity* with *type identity*. The external map tracks which specific object instance came from where (not which *type* in the MRO provided a value).

Consider:

```
class A:
    x = 1

class B(A):
    pass # Inherits x from A

b = B()
print(b.x) # Prints 1. Which type provided this?
```

An external provenance map could record `provenance_map[id(b)] = B`. But this doesn’t answer the question “which type in B’s MRO provided `x`?”. The answer is `A`, and this requires MRO traversal, which requires the Bases axis.

Formal statement: Let `ExternalMap : ObjectId → SourceType` be any external provenance map. Then:

`ExternalMap` cannot answer: “Which type in `MRO(type(obj))` provided at

Proof. The question asks about MRO position. MRO is derived from Bases. `ExternalMap` has no access to Bases (it maps object IDs to types, not types to MRO positions). Therefore `ExternalMap` cannot answer MRO-position queries. ■

The deeper point: Provenance is not about “where did this object come from?” It’s about “where did this *value* come from in the inheritance hierarchy?” The latter requires MRO, which requires Bases, which interface-only observation discards.

M. Abstract Model Lean Formalization

The abstract class system model (Section 2.4) is formalized in Lean 4 with complete proofs (no sorry placeholders):

```
-- The two axes of a class system
-- NOTE: "Name" (N) is NOT an independent axis, it is metadata stored in the type
-- the namespace (S axis) as __name__. It contributes no typing capability.
-- The minimal model is (B, S).
inductive Axis where
| Bases      -- B: inheritance hierarchy
| Namespace   -- S: attribute declarations
deriving DecidableEq, Repr

-- A typing discipline is characterized by which axes it inspects
abbrev AxisSet := List Axis

-- Canonical axis sets
def shapeAxes : AxisSet := [.Namespace] -- S-only: only interface, no sys interface
def nominalAxes : AxisSet := [.Bases, .Namespace] (B, S): full nominal

-- Unified capability (combines typing and architecture domains)
inductive UnifiedCapability where
| interfaceCheck    -- Check interface specifications_nominal_exceeds_shape :
| identity          -- Type identity           exists c in axisSetCapabilities nominalAxes,
| provenance        -- Type provenance         cnotin axisSetCapabilities shapeAxes := by
| enumeration       -- Subtype enumerationuse UnifiedCapability.provenance
| conflictResolution -- MRO-based resolutiononstructor
deriving DecidableEq, Repr

-- Capabilities enabled by each axis
def axisCapabilities (a : Axis) : List UnifiedCapability = METATHEOREM: Combined strict dominance
match a with
| .Bases => [.identity, .provenance, .enumeration]
| .Namespace => [.interfaceCheck]
```

-- Capabilities of an axis set = union of each axis's capabilities
This formalizes Theorem 2.15: using more axes provides strictly more capabilities. The proofs are complete and compile without any sorry placeholders.

Definition 6.3a (Axis Projection — Lean). A type over axis set A is a dependent tuple of axis values:

```
-- Type as projection: for each axis in A, provide possible capabilities in the system
def AxisProjection (A : List Axis) := (a : Axis) -- Inductive Capability Axises here
| interfaceCheck    -- "Does x have method m?"
| valueAccess       -- "What is the name of type x?"
| methodInvocation  -- "Call x.m()"
| provenance        -- "Which type provided the value x?"
| identity          -- "Is x an instance of T?"
| enumeration       -- "What are all subtypes of T?"
| conflictResolution -- "Which definition wins?"
```

-- Concrete Typ is a 2-tuple: (namespace, base\$)typeNaming

structure Typ where
 ns : Finset AttrName
 bs : List Typ

Theorem 6.3b (Isomorphism Theorem — Lean). The concrete type Typ is isomorphic to the 2-axis projection:

-- The Isomorphism: Typ <-> AxisProjection {Basing DecidableEq, Repr}

noncomputable def Typ.equivProjection : Typ <-> AxisProjection canonicalAxes where

toFun := Typ.toProjection -- Capabilities that require the Bases axis
 invFun := Typ.fromProjection -- basesRequiredCapabilities : List Capability :=

left_inv := Typ.projection_roundtrip -- [fromProjection.onIdentityProjection, .conflictResolution]
 right_inv := Typ.projection_roundtrip_inv -- toProjection . fromProjection = id

-- Capabilities that do NOT require Bases (only need

```
-- For any axis set A, GenericTyp A IS AxisProjection
theorem n_axis_types_are_projections (A : List Axis)
  GenericTyp A = AxisProjection A := rfl
```

This formalizes Definition 2.10 (Typing Disciplines as Axis Projections): a typing discipline using axis set A has exactly the information contained in the A -projection of the full type.

Theorem 6.4 (Axis Lattice — Lean). Shape capabilities are a strict subset of nominal capabilities:

```
-- THEOREM: Shape axes subset Nominal axes (specified by shapeAxes)
theorem axis_shape_subset_nominal :
  forall c in axisSetCapabilities shapeAxes,
  c in axisSetCapabilities nominalAxes := by
  have h_shape : axisSetCapabilities shapeAxes = []
  have h_nominal : UnifiedCapability.interfaceCheck
  rw [h_shape] at hc
  exact hc
```

-- S-only: only interface, no sys interface (intentional observation)

def nominalAxes : AxisSet := [.Bases, .Namespace] (B, S): full nominal

exact h_nominal

-- THEOREM: Nominal has capabilities Shape lacks

theorem shape_exceeds_shape :
 exists c in axisSetCapabilities nominalAxes,
 cnotin axisSetCapabilities shapeAxes := by

* decide -- provenance in nominalAxes capabilities
 * decide -- provenance notin shapeAxes capabilities

theorem lattice_dominance :
 exists c in axisSetCapabilities nominalAxes,
 <axis_shape_subset_nominal, axis_nominal_exceeds_shape>

def axisSetCapabilities (axes : AxisSet) : List UnifiedCapability = METATHEOREM: Combined strict dominance

axes.flatMap axisCapabilities |>.eraseDups

Theorem 6.11 (Capability Completeness — Lean). The Bases axis provides exactly four capabilities, no more:

```

def nonBasesCapabilities : List Capability := Protocol runtime observer (shape-only). We also
  [.interfaceCheck, .typeNaming, .valueAccess] formalized that the [restricted] Protocol/instance observer
-- THEOREM: Bases capabilities are exactly that checks only for required members. The predicate
theorem bases_capabilities_complete : {protoCheck, ignoresProtocolIdentity} and isProvided[Resolut
  forall c : Capability, shape-respecting (protoCheck_in_shapeQuerySet
  (c in basesRequiredCapabilities <=> c = .provenance \vee c = .identity) \vee
  intro c in abstract_class_system.lean), so two protocols
constructor constructor Distinguishing them requires adding .isObservable) := by
  * intro h
    simp [basesRequiredCapabilities] at h
    exact h
  * intro h
    simp [basesRequiredCapabilities]
    exact h
discriminator (brand/tag/nominality), i.e., moving to another axis.

-- THEOREM: Non-Bases capabilities are exactly they{yield rfbBaseCheck, typeNamingThroughNamespace, method
theorem non_bases_capabilities_complete : python_instantiation.lean.
  forall c : Capability,
  (c in nonBasesCapabilities <=>
  c = .interfaceCheck \vee c = .typeNaming \vee c = .valueAccess \vee c = .methodInvocation) :=
  intro c
  constructor
  * intro h
    simp [nonBasesCapabilities] at h
    exact h
  * intro h
    simp [nonBasesCapabilities]
    exact h

-- THEOREM: Every capability is in exactly one category (partition)
theorem capability_partition :
  forall c : Capability,
  (c in basesRequiredCapabilities \vee c in nonBasesCapabilities) /\ ~(c in basesRequiredCapabilities \wedge c in nonBasesCapabilities) := by
  intro c
  cases c < simp [basesRequiredCapabilities, nonBasesCapabilities]

-- THEOREM: |basesRequiredCapabilities| = 4 (exactly four capabilities)
theorem bases_capabilities_count :
  basesRequiredCapabilities.length = 4 := by rfl

```

This formalizes Theorem 2.17 (Capability Completeness):
the capability set \mathcal{C}_B is **exactly** four elements, proven by
exhaustive enumeration with machine-checked partition. The
capability_partition theorem proves that every capa-
bility falls into exactly one category (Bases-required or not)
with no overlap and no gaps.

Scope as observational quotient. We model “scope” as
a set of allowed observers $\text{Obs} \subseteq (W \rightarrow O)$ and de-
fine observational equivalence $x \approx y : \iff \forall f \in$
 $\text{Obs}, f(x) = f(y)$. The induced quotient W/\approx is the
canonical object for that scope, and every in-scope ob-
server factors through it (see `observer_factors` in
`abstract_class_system.lean`). Once the observer set
is fixed, no argument can appeal to information outside that
quotient; adding a new observable is literally expanding Obs .

All Python object-model observables factor through axes. In the Python instantiation we prove that core runtime discriminators are functions of (B, S) : metaclass selection depends only on bases (metaclass_depends_on_bases); attribute presence and dispatch depend only on the namespace (getattr_depends_on_ns); together

they{yield rfbBaseCheck, typeNamingThroughNamespace, method instantiation.lean}.