The Complexity of Decision-Relevant Uncertainty: Why Identifying What Matters Is Harder Than Knowing Everything

Tristan Simas
McGill University
tristan.simas@mail.mcgill.ca

January 3, 2026

Abstract

Engineers routinely include irrelevant information in their models. Climate scientists model atmospheric chemistry when predicting regional temperatures. Financial analysts track hundreds of indicators when making portfolio decisions. Software architects specify dozens of configuration parameters when only a handful affect outcomes.

This paper proves that such *over-modeling* is not laziness—it is computationally rational. Identifying precisely which variables are "decision-relevant" is coNP-complete [6, 14], finding the *minimum* set of relevant variables is coNP-complete, and a fixed-coordinate "anchor" version is Σ_2^P-complete [29]. These results formalize a fundamental insight:

Determining what you need to know is harder than knowing everything.

We introduce the *decision quotient*—a measure of decision-relevant complexity—and prove a complexity dichotomy: checking sufficiency is polynomial when the minimal sufficient set has logarithmic size, but exponential when it has linear size. We identify tractable subcases (bounded actions, separable utilities, tree-structured dependencies) that admit polynomial algorithms.

A major practical consequence is the *Simplicity Tax Theorem*: using a simple tool for a complex problem is necessarily harder than using a tool matched to the problem's complexity. When a tool lacks native support for required dimensions, users must supply that information externally at every use site. The "simpler" tool creates more total work, not less. This overturns the common intuition that "simpler is always better"—simplicity is only a virtue when the problem is also simple.

These are ceiling results: The complexity characterizations are exact (both upper and lower bounds). The theorems quantify universally over all problem instances ($\forall$), not probabilistically ($\mu = 1$). The dichotomy is complete—no intermediate cases exist under standard assumptions. The tractability conditions are maximal—relaxing any yields hardness. No stronger complexity claims are possible within classical complexity theory.

All results are machine-checked in Lean 4 [8] ($\sim$5,000 lines across 33 files, 200+ theorems). The Lean formalization proves: (1) polynomial-time reduction composition; (2) correctness of the TAUTOLOGY and $\exists\forall$-SAT reduction mappings; (3) equivalence of sufficiency checking with coNP/Σ_2^P-complete problems under standard encodings; (4) the Simplicity Tax Theorem including conservation, dominance, and the amortization threshold. Complexity classifications (coNP-complete, Σ_2^P-complete) are derived by combining these machine-checked results with the well-known complexity of TAUTOLOGY and $\exists\forall$-SAT.

1 Introduction

This paper establishes a fundamental limit on rational decision-making under uncertainty:

> **Determining what you need to know is harder than knowing everything.**

This is not metaphor. It is a theorem. Specifically: given a decision problem with n dimensions of uncertainty, *checking* whether a subset of dimensions suffices for optimal action is coNP-complete. *Finding* the minimal sufficient subset is coNP-complete. These results hold universally—for any decision problem with coordinate structure.

The implications are immediate and far-reaching. Engineers who include "irrelevant" information in their models are not exhibiting poor discipline. They are responding optimally to a computational constraint that admits no workaround. Climate scientists modeling atmospheric chemistry, financial analysts tracking hundreds of indicators, software architects specifying dozens of parameters—all are exhibiting computationally rational behavior. The alternative (identifying precisely which variables matter) requires solving coNP-complete problems.

1.1 The Core Problem

Consider a decision problem with actions A and states $S = X_1 \times \cdots \times X_n$ (a product of n coordinate spaces). For each state $s \in S$, some subset $\mathrm{Opt}(s) \subseteq A$ of actions are optimal. The fundamental question is:

> **Which coordinates are sufficient to determine the optimal action?**

A coordinate set $I \subseteq \{1, \ldots, n\}$ is *sufficient* if knowing only the coordinates in I determines the optimal action set:
$$s_I = s'_I \implies \mathrm{Opt}(s) = \mathrm{Opt}(s')$$
where s_I denotes the projection of state s onto coordinates in I.

1.2 Main Results

This paper proves four main theorems:

1. **Theorem 3.6 (Sufficiency Checking is coNP-complete):** Given a decision problem and coordinate set I, determining whether I is sufficient is coNP-complete [6, 14].

2. **Theorem 3.7 (Minimum Sufficiency is coNP-complete):** Finding the minimum sufficient coordinate set is coNP-complete. (The problem is trivially in Σ_2^{P} by structure, but collapses to coNP because sufficiency equals "superset of relevant coordinates.")

3. **Theorem 4.1 (Complexity Dichotomy):** Sufficiency checking exhibits a dichotomy:

   - If the minimal sufficient set has size $O(\log |S|)$, checking is polynomial
   - If the minimal sufficient set has size $\Omega(n)$, checking requires exponential time [13].

4. **Theorem 5.1 (Tractable Subcases):** Sufficiency checking is polynomial-time for:

- Bounded action sets ($|A| \leq k$ for constant k)
- Separable utility functions ($u(a, s) = f(a) + g(s)$)
- Tree-structured coordinate dependencies

1.3   The Foundational Principle

The core result transcends the specific application domain:

> **For any agent facing structured uncertainty, identifying the relevant dimensions of uncertainty is computationally harder than simply observing all dimensions.**

This applies to:

- **Machine learning:** Feature selection is intractable in general

- **Economics:** Identifying relevant market factors is intractable

- **Scientific modeling:** Determining which variables matter is intractable

- **Software engineering:** Configuration minimization is intractable

The ubiquity of over-modeling, over-parameterization, and "include everything" strategies across domains is not coincidence. It is the universal rational response to a universal computational constraint.

1.4   Connection to Prior Papers

This paper completes the theoretical foundation established in Papers 1–3:

- **Paper 1 (Typing):** Showed nominal typing dominates structural typing

- **Paper 2 (SSOT):** Showed single source of truth minimizes modification complexity

- **Paper 3 (Leverage):** Unified both as leverage maximization

Paper 4's contribution: Proves that *identifying* which architectural decisions matter is itself computationally hard. This explains why leverage maximization (Paper 3) uses heuristics rather than optimal algorithms—and why this is not a deficiency but a mathematical necessity.

1.5   The Simplicity Tax: A Major Practical Consequence

Beyond the complexity-theoretic results, this paper develops a foundational practical principle: the *Simplicity Tax Theorem*.

The common intuition "simpler is better" is context-dependent. When a problem has intrinsic complexity (many required axes), using a "simple" tool (few native axes) forces the complexity elsewhere—to every use site. This paper proves:

> **Using a simple tool for a complex problem is necessarily harder than using a tool matched to the problem's complexity.**

Specifically: if a tool covers k of n required axes, the remaining $n-k$ axes become the *simplicity tax*, paid at *every use site*. For m use sites, total external work is $(n-k) \times m$. A complete tool (covering all axes) pays zero tax.

This overturns simplicity-as-virtue folklore. Preferring "simple" tools for complex problems is not wisdom—it is a failure to account for distributed costs. True sophistication is matching tool complexity to problem complexity.

Section 8 develops this theorem formally. All results are machine-checked in Lean 4.

1.6 Paper Structure

Section 2 establishes formal foundations: decision problems, coordinate spaces, sufficiency. Section 3 proves hardness results with complete reductions. Section 4 develops the complexity dichotomy. Section 5 presents tractable special cases. Section 7 discusses implications for software architecture, including hardness distribution. Section 8 develops the Simplicity Tax Theorem as a major practical consequence. Section 9 surveys related work. Appendix A contains Lean proof listings. Appendix B addresses anticipated objections.

1.7 Anticipated Objections

Before proceeding, we address objections readers are likely forming. Each is refuted in detail in Appendix B; here we summarize the key points.

"coNP-completeness doesn't mean intractable—there might be good heuristics." Correct, but this strengthens our thesis. The point is not that practitioners cannot find useful approximations, but that *optimal* dimension selection is provably hard. The prevalence of heuristics (feature selection in ML, sensitivity analysis in economics) is itself evidence of the computational barrier.

"Real decision problems don't have clean coordinate structure." The coordinate structure assumption is weaker than it appears. Any finite state space can be encoded with binary coordinates; the hardness results apply to this encoding. More structured representations make the problem *easier*, not harder—so hardness for structured problems implies hardness for general ones.

"The reduction from SAT is artificial." All coNP-completeness proofs use reductions. The reduction demonstrates that SAT instances can be encoded as sufficiency-checking problems while preserving computational structure. This is standard complexity theory methodology [6, 14]. The claim is not that practitioners encounter SAT problems, but that sufficiency checking is at least as hard as SAT.

"The tractable subcases are too restrictive to be useful." The tractable subcases (bounded actions, separable utility, tree structure) characterize *when* dimension selection becomes feasible. Many real problems fall into these categories. The dichotomy theorem (Theorem 4.1) precisely identifies the boundary between tractable and intractable.

"This just formalizes the obvious—of course feature selection is hard." The contribution is making "obvious" precise. Prior work established heuristic hardness for specific domains (ML feature selection, economic factor identification). We prove a *universal* result that applies to *any* decision problem with coordinate structure. This unification is the theoretical contribution.

4

If you have an objection not listed above, check Appendix B (12 objections addressed, including 4 specific to the Simplicity Tax) before concluding it has not been considered.

2 Formal Foundations

We formalize decision problems with coordinate structure, sufficiency of coordinate sets, and the decision quotient, drawing on classical decision theory [24, 22].

2.1 Decision Problems with Coordinate Structure

Definition 2.1 (Decision Problem). A *decision problem with coordinate structure* is a tuple $\mathcal{D} = (A, X_1, \ldots, X_n, U)$ where:

- A is a finite set of *actions* (alternatives)

- $X_1, \ldots, X_n$ are finite *coordinate spaces*

- $S = X_1 \times \cdots \times X_n$ is the *state space*

- $U : A \times S \to \mathbb{Q}$ is the *utility function*

Definition 2.2 (Projection). For state $s = (s_1, \ldots, s_n) \in S$ and coordinate set $I \subseteq \{1, \ldots, n\}$:

$$s_I := (s_i)_{i \in I}$$

is the *projection* of s onto coordinates in I.

Definition 2.3 (Optimizer Map). For state $s \in S$, the *optimal action set* is:

$$\mathrm{Opt}(s) := \arg\max_{a \in A} U(a, s) = \{a \in A : U(a, s) = \max_{a' \in A} U(a', s)\}$$

2.2 Sufficiency and Relevance

Definition 2.4 (Sufficient Coordinate Set). A coordinate set $I \subseteq \{1, \ldots, n\}$ is *sufficient* for decision problem $\mathcal{D}$ if:

$$\forall s, s' \in S : \quad s_I = s'_I \implies \mathrm{Opt}(s) = \mathrm{Opt}(s')$$

Equivalently, the optimal action depends only on coordinates in I.

Definition 2.5 (Minimal Sufficient Set). A sufficient set I is *minimal* if no proper subset $I' \subsetneq I$ is sufficient.

Definition 2.6 (Relevant Coordinate). Coordinate i is *relevant* if it belongs to some minimal sufficient set.

Example 2.7 (Weather Decision). Consider deciding whether to carry an umbrella:

- Actions: $A = \{\text{carry}, \text{don't carry}\}$

- Coordinates: $X_1 = \{\text{rain}, \text{no rain}\}$, $X_2 = \{\text{hot}, \text{cold}\}$, $X_3 = \{\text{Monday}, \ldots, \text{Sunday}\}$

- Utility: $U(\text{carry}, s) = -1 + 3 \cdot \mathbf{1}[s_1 = \text{rain}]$, $U(\text{don't carry}, s) = -2 \cdot \mathbf{1}[s_1 = \text{rain}]$

The minimal sufficient set is $I = \{1\}$ (only rain forecast matters). Coordinates 2 and 3 (temperature, day of week) are irrelevant.

2.3 The Decision Quotient

Definition 2.8 (Decision Equivalence)**.** For coordinate set I, states s, s' are *I-equivalent* (written $s \sim_I s'$) if $s_I = s'_I$.

Definition 2.9 (Decision Quotient)**.** The *decision quotient* for state s under coordinate set I is:

$$\mathrm{DQ}_I(s) = \frac{|\{a \in A : a \in \mathrm{Opt}(s') \text{ for some } s' \sim_I s\}|}{|A|}$$

This measures the fraction of actions that *could* be optimal given only the information in I.

Proposition 2.10 (Sufficiency Characterization)**.** *Coordinate set I is sufficient if and only if $DQ_I(s) = |Opt(s)|/|A|$ for all $s \in S$.*

Proof. If I is sufficient, then $s \sim_I s' \implies \mathrm{Opt}(s) = \mathrm{Opt}(s')$, so the set of actions optimal for some $s' \sim_I s$ is exactly $\mathrm{Opt}(s)$.

Conversely, if the condition holds, then for any $s \sim_I s'$, the optimal actions form the same set (since $\mathrm{DQ}_I(s) = \mathrm{DQ}_I(s')$ and both equal the relative size of the common optimal set). ■

3 Computational Complexity of Decision-Relevant Uncertainty

This section establishes the computational complexity of determining which state coordinates are decision-relevant. We prove three main results:

1. **SUFFICIENCY-CHECK** is coNP-complete

2. **MINIMUM-SUFFICIENT-SET** is coNP-complete (the Σ_2^P structure collapses)

3. **ANCHOR-SUFFICIENCY** (fixed coordinates) is Σ_2^P-complete

These results sit beyond NP-completeness and formally explain why engineers default to over-modeling: finding the minimal set of decision-relevant factors is computationally intractable.

3.1 Problem Definitions

Definition 3.1 (Decision Problem Encoding)**.** A *decision problem instance* is a tuple (A, n, U) where:

- A is a finite set of alternatives

- n is the number of state coordinates, with state space $S = \{0, 1\}^n$

- $U : A \times S \to \mathbb{Q}$ is the utility function, given as a Boolean circuit

Definition 3.2 (Optimizer Map)**.** For state $s \in S$, define:

$$\mathrm{Opt}(s) := \arg\max_{a \in A} U(a, s)$$

Definition 3.3 (Sufficient Coordinate Set)**.** A coordinate set $I \subseteq \{1, \dots, n\}$ is *sufficient* if:

$$\forall s, s' \in S : \quad s_I = s'_I \implies \mathrm{Opt}(s) = \mathrm{Opt}(s')$$

where s_I denotes the projection of s onto coordinates in I.

Problem 3.4 (SUFFICIENCY-CHECK). **Input:** Decision problem (A, n, U) and coordinate set $I \subseteq \{1, \ldots, n\}$
Question: Is I sufficient?

Problem 3.5 (MINIMUM-SUFFICIENT-SET). **Input:** Decision problem (A, n, U) and integer k
Question: Does there exist a sufficient set I with $|I| \leq k$?

3.2   Hardness of SUFFICIENCY-CHECK

Theorem 3.6 (coNP-completeness of SUFFICIENCY-CHECK). *SUFFICIENCY-CHECK is coNP-complete [6, 14].*

Proof. **Membership in coNP:** The complementary problem INSUFFICIENCY is in NP. Given (A, n, U, I), a witness for insufficiency is a pair (s, s') such that:

1. $s_I = s'_I$ (verifiable in polynomial time)

2. $\mathrm{Opt}(s) \neq \mathrm{Opt}(s')$ (verifiable by evaluating U on all alternatives)

   coNP-hardness: We reduce from TAUTOLOGY.
   Given Boolean formula $\varphi(x_1, \ldots, x_n)$, construct a decision problem with:

   - Alternatives: $A = \{\mathrm{accept}, \mathrm{reject}\}$

   - State space: $S = \{\mathrm{reference}\} \cup \{0, 1\}^n$

   - Utility:

$$\begin{aligned}
U(\mathrm{accept}, \mathrm{reference}) &= 1 \\
U(\mathrm{reject}, \mathrm{reference}) &= 0 \\
U(\mathrm{accept}, a) &= \varphi(a) \\
U(\mathrm{reject}, a) &= 0 \quad \text{for assignments } a \in \{0, 1\}^n
\end{aligned}$$

   - Query set: $I = \emptyset$

   Claim: $I = \emptyset$ is sufficient $\iff$ φ is a tautology.
   ($\Rightarrow$) Suppose I is sufficient. Then $\mathrm{Opt}(s)$ is constant over all states. Since $U(\mathrm{accept}, a) = \varphi(a)$ and $U(\mathrm{reject}, a) = 0$:

   - $\mathrm{Opt}(a) = \mathrm{accept}$ when $\varphi(a) = 1$

   - $\mathrm{Opt}(a) = \{\mathrm{accept}, \mathrm{reject}\}$ when $\varphi(a) = 0$

For Opt to be constant, $\varphi(a)$ must be true for all assignments a; hence φ is a tautology.
   ($\Leftarrow$) If φ is a tautology, then $U(\mathrm{accept}, a) = 1 > 0 = U(\mathrm{reject}, a)$ for all assignments a. Thus $\mathrm{Opt}(s) = \{\mathrm{accept}\}$ for all states s, making $I = \emptyset$ sufficient.   ∎

3.3 Complexity of MINIMUM-SUFFICIENT-SET

Theorem 3.7 (MINIMUM-SUFFICIENT-SET is coNP-complete). *MINIMUM-SUFFICIENT-SET is coNP-complete.*

Proof. **Structural observation:** The $\exists\forall$ quantifier pattern suggests Σ_2^P:

$$\exists I \, (|I| \leq k) \, \forall s, s' \in S : \quad s_I = s'_I \implies \mathrm{Opt}(s) = \mathrm{Opt}(s')$$

However, this collapses because sufficiency has a simple characterization.

Key lemma: A coordinate set I is sufficient if and only if I contains all relevant coordinates (proven formally as `sufficient_contains_relevant` in Lean):

$$\mathrm{sufficient}(I) \iff \mathrm{Relevant} \subseteq I$$

where $\mathrm{Relevant} = \{i : \exists s, s'. \ s \text{ differs from } s' \text{ only at } i \text{ and } \mathrm{Opt}(s) \neq \mathrm{Opt}(s')\}$.

Consequence: The minimum sufficient set is exactly the set of relevant coordinates. Thus MINIMUM-SUFFICIENT-SET asks: "Is the number of relevant coordinates at most k?"

coNP membership: A witness that the answer is NO is a set of $k+1$ coordinates, each proven relevant (by exhibiting s, s' pairs). Verification is polynomial.

coNP-hardness: The $k = 0$ case asks whether no coordinates are relevant, i.e., whether $\emptyset$ is sufficient. This is exactly SUFFICIENCY-CHECK, which is coNP-complete by Theorem 3.6. ∎

3.4 Anchor Sufficiency (Fixed Coordinates)

We also formalize a strengthened variant that fixes the coordinate set and asks whether there exists an *assignment* to those coordinates that makes the optimal action constant on the induced subcube.

Problem 3.8 (ANCHOR-SUFFICIENCY). **Input:** Decision problem (A, n, U) and fixed coordinate set $I \subseteq \{1, \ldots, n\}$
Question: Does there exist an assignment α to I such that $\mathrm{Opt}(s)$ is constant for all states s with $s_I = \alpha$?

Theorem 3.9 (ANCHOR-SUFFICIENCY is Σ_2^P-complete). *ANCHOR-SUFFICIENCY is Σ_2^P-complete [29] (already for Boolean coordinate spaces).*

Proof. **Membership in Σ_2^P:** The problem has the form

$$\exists \alpha \, \forall s \in S : (s_I = \alpha) \implies \mathrm{Opt}(s) = \mathrm{Opt}(s_\alpha),$$

which is an $\exists\forall$ pattern.

Σ_2^P-hardness: Reduce from $\exists\forall$-SAT. Given $\exists x \, \forall y \, \varphi(x, y)$ with $x \in \{0, 1\}^k$ and $y \in \{0, 1\}^m$, if $m = 0$ we first pad with a dummy universal variable (satisfiability is preserved), construct a decision problem with:

- Actions $A = \{\mathrm{YES}, \mathrm{NO}\}$

- State space $S = \{0, 1\}^{k+m}$ representing (x, y)

- Utility
$$U(\mathrm{YES}, (x, y)) = \begin{cases} 2 & \text{if } \varphi(x, y) = 1 \\ 0 & \text{otherwise} \end{cases} \qquad U(\mathrm{NO}, (x, y)) = \begin{cases} 1 & \text{if } y = 0^m \\ 0 & \text{otherwise} \end{cases}$$

- Fixed coordinate set $I = $ the x-coordinates.

If $\exists x^\star \, \forall y \, \varphi(x^\star, y) = 1$, then for any y we have $U(\text{YES}) = 2$ and $U(\text{NO}) \leq 1$, so $\text{Opt}(x^\star, y) = \{\text{YES}\}$ is constant. Conversely, if $\varphi(x, y)$ is false for some y, then either $y = 0^m$ (where NO is optimal) or $y \neq 0^m$ (where YES and NO tie), so the optimal set varies across y and the subcube is not constant. Thus an anchor assignment exists iff the $\exists \forall$-SAT instance is true. $\blacksquare$

3.5 Tractable Subcases

Despite the general hardness, several natural subcases admit efficient algorithms:

Proposition 3.10 (Small State Space). *When $|S|$ is polynomial in the input size (i.e., explicitly enumerable), MINIMUM-SUFFICIENT-SET is solvable in polynomial time.*

Proof. Compute $\text{Opt}(s)$ for all $s \in S$. The minimum sufficient set is exactly the set of coordinates that "matter" for the resulting function, computable by standard techniques. $\blacksquare$

Proposition 3.11 (Linear Utility). *When $U(a, s) = w_a \cdot s$ for weight vectors $w_a \in \mathbb{Q}^n$, MINIMUM-SUFFICIENT-SET reduces to identifying coordinates where weight vectors differ.*

3.6 Implications

Corollary 3.12 (Why Over-Modeling Is Rational). *Finding the minimal set of decision-relevant factors is coNP-complete.* Even *verifying that a proposed set is sufficient is coNP-complete.*
This formally explains the engineering phenomenon:

1. *It's computationally easier to model everything than to find the minimum*

2. *"Which unknowns matter?" is a hard question, not a lazy one to avoid*

3. *Bounded scenario analysis (small $\hat{S}$) makes the problem tractable*

This connects to the pentalogy's leverage framework: the "epistemic budget" for deciding what to model is itself a computationally constrained resource.

3.7 Remark: The Collapse to coNP

Early analysis of MINIMUM-SUFFICIENT-SET focused on the apparent $\exists \forall$ quantifier structure, which suggested a Σ_2^P-complete result. We initially explored certificate-size lower bounds for the complement, attempting to show MINIMUM-SUFFICIENT-SET was unlikely to be in coNP.
However, the key insight—formalized as `sufficient_contains_relevant`—is that sufficiency has a simple characterization: a set is sufficient iff it contains all relevant coordinates. This collapses the $\exists \forall$ structure because:

- The minimum sufficient set is *exactly* the relevant coordinate set

- Checking relevance is in coNP (witness: two states differing only at that coordinate with different optimal sets)

- Counting relevant coordinates is also in coNP

This collapse explains why ANCHOR-SUFFICIENCY retains its Σ_2^P-completeness: fixing coordinates and asking for an assignment that works is a genuinely different question. The "for all suffixes" quantifier cannot be collapsed when the anchor assignment is part of the existential choice.

4   Complexity Dichotomy

The hardness results of Section 3 apply to worst-case instances. This section develops a more nuanced picture: a *dichotomy theorem* showing that problem difficulty depends on the size of the minimal sufficient set.

Theorem 4.1 (Complexity Dichotomy). *Let $\mathcal{D} = (A, X_1, \ldots, X_n, U)$ be a decision problem with $|S| = N$ states. Let k^* be the size of the minimal sufficient set.*

1. ***Logarithmic case:*** *If $k^* = O(\log N)$, then SUFFICIENCY-CHECK is solvable in polynomial time.*

2. ***Linear case:*** *If $k^* = \Omega(n)$, then SUFFICIENCY-CHECK requires time $\Omega(2^{n/c})$ for some constant $c > 0$ (assuming ETH).*

Proof. **Part 1 (Logarithmic case):** If $k^* = O(\log N)$, then the number of distinct projections $|S_{I^*}|$ is at most $2^{k^*} = O(N^c)$ for some constant c. We can enumerate all projections and verify sufficiency in polynomial time.

   Part 2 (Linear case): We establish this via an explicit reduction chain from the Exponential Time Hypothesis. ■

4.1   The ETH Reduction Chain

The lower bound in Part 2 of Theorem 4.1 follows from a chain of reductions originating in the Exponential Time Hypothesis. We make this chain explicit.

Definition 4.2 (Exponential Time Hypothesis (ETH)). There exists a constant $\delta > 0$ such that 3-SAT on n variables cannot be solved in time $O(2^{\delta n})$ [13].

The chain proceeds as follows:

1. **ETH $\Rightarrow$ 3-SAT requires $2^{\Omega(n)}$:** This is the definition of ETH.

2. **3-SAT $\leq_p$ TAUTOLOGY:** Given 3-SAT formula $\varphi(x_1, \ldots, x_n)$, define $\psi = \neg\varphi$. Then φ is satisfiable iff ψ is not a tautology. This is a linear-time reduction preserving the number of variables.

3. **TAUTOLOGY requires $2^{\Omega(n)}$ (under ETH):** By the contrapositive of step 2, if TAUTOLOGY could be solved in $o(2^{\delta n})$ time, then 3-SAT could be solved in $o(2^{\delta n})$ time, contradicting ETH.

4. **TAUTOLOGY $\leq_p$ SUFFICIENCY-CHECK:** This is Theorem 3.6. Given formula $\varphi(x_1, \ldots, x_n)$, we construct a decision problem where:

   - The empty set $I = \emptyset$ is sufficient iff φ is a tautology
   - When φ is not a tautology, all n coordinates are relevant

   The reduction is polynomial-time and preserves the number of coordinates.

5. **SUFFICIENCY-CHECK requires $2^{\Omega(n)}$ (under ETH):** Combining steps 3 and 4: if SUFFICIENCY-CHECK could be solved in $o(2^{\delta n/c})$ time for some constant c, then TAUTOLOGY (and hence 3-SAT) could be solved in subexponential time, contradicting ETH.

Proposition 4.3 (Tight Constant). *The reduction in Theorem 3.6 preserves the number of variables exactly: an n-variable formula yields an n-coordinate decision problem. Therefore, the constant c in the $2^{n/c}$ lower bound equals 1:*

$$SUFFICIENCY\text{-}CHECK \text{ requires time } \Omega(2^{\delta n}) \text{ under ETH}$$

where δ is the ETH constant for 3-SAT.

Proof. The TAUTOLOGY reduction (Theorem 3.6) constructs:

- State space $S = \{\text{ref}\} \cup \{0,1\}^n$ with $n+1$ coordinates (one extra for the reference state)

- Query set $I = \emptyset$

When φ has n variables, the constructed problem has $n+1$ coordinates. The asymptotic lower bound is $2^{\Omega(n)}$ with the same constant δ from ETH. $\blacksquare$

4.2 Phase Transition

Corollary 4.4 (Phase Transition). *There exists a threshold $\tau \in (0,1)$ such that:*

- *If $k^*/n < \tau$, SUFFICIENCY-CHECK is "easy" (polynomial in N)*

- *If $k^*/n > \tau$, SUFFICIENCY-CHECK is "hard" (exponential in n)*

Proof. The logarithmic case (Part 1 of Theorem 4.1) gives polynomial time when $k^* = O(\log N) = O(\log 2^n) = O(n)$. More precisely, when $k^* \leq c \log N$ for constant c, the algorithm runs in time $O(N^c \cdot \text{poly}(n))$.

The linear case (Part 2) gives exponential time when $k^* = \Omega(n)$.

The threshold τ is implicitly defined by where the polynomial bound $2^{k^*} = N^{k^*/\log N}$ transitions from polynomial to superpolynomial in N. This occurs when $k^*/\log N$ exceeds any constant, i.e., when $k^* = \omega(\log N)$.

For Boolean coordinate spaces ($N = 2^n$), the threshold is $\tau = 0$: any $k^* = \omega(\log n)$ yields superpolynomial complexity, while $k^* = O(\log n)$ is tractable. $\blacksquare$

Remark 4.5 (Sharpness of Dichotomy). The dichotomy is asymptotically tight under ETH. There are no "intermediate" cases:

- **Below threshold:** Polynomial (by enumeration)

- **At threshold:** Quasipolynomial ($n^{O(\log n)}$)

- **Above threshold:** Exponential (by ETH)

The quasipolynomial regime at the threshold is measure-zero; almost all instances are either clearly tractable or clearly intractable.

This dichotomy explains why some domains admit tractable model selection (few relevant variables) while others require heuristics (many relevant variables). The ETH reduction chain makes precise what "hard" means: not merely coNP-complete, but requiring $2^{\Omega(n)}$ time under widely-believed complexity assumptions.

5   Tractable Special Cases

Despite the general hardness, several natural problem classes admit polynomial-time algorithms.

Theorem 5.1 (Tractable Subcases). *SUFFICIENCY-CHECK is polynomial-time solvable for:*

1. ***Bounded actions:*** $|A| \leq k$ *for constant* k

2. ***Separable utility:*** $U(a, s) = f(a) + g(s)$

3. ***Tree-structured dependencies:*** *Coordinates form a tree where each coordinate depends only on its ancestors*

5.1   Bounded Actions

Proof of Part 1. With $|A| = k$ constant, the optimizer map $\text{Opt} : S \to 2^A$ has at most 2^k distinct values. For each pair of distinct optimizer values, we can identify the coordinates that distinguish them. The union of these distinguishing coordinates forms a sufficient set.

The algorithm:

1. Sample states to identify distinct optimizer values (polynomial samples suffice with high probability)

2. For each pair of optimizer values, find distinguishing coordinates

3. Return the union of distinguishing coordinates

This runs in time $O(|S| \cdot k^2)$ which is polynomial when k is constant. ∎

5.2   Separable Utility

Proof of Part 2. If $U(a, s) = f(a) + g(s)$, then:

$$\text{Opt}(s) = \arg\max_{a \in A}[f(a) + g(s)] = \arg\max_{a \in A} f(a)$$

The optimal action is independent of the state! Thus $I = \emptyset$ is always sufficient. ∎

5.3   Tree-Structured Dependencies

Proof of Part 3. When coordinates form a tree, we can use dynamic programming. For each node i, compute the set of optimizer values achievable in the subtree rooted at i. A coordinate is relevant if and only if different values at that coordinate lead to different optimizer values in its subtree. This approach is analogous to inference in probabilistic graphical models [20, 15].

The algorithm runs in time $O(n \cdot |A|^2)$ by processing the tree bottom-up. ∎

5.4   Practical Implications

These tractable cases correspond to common modeling scenarios:

- **Bounded actions:** Most real decisions have few alternatives (buy/sell/hold, approve/reject, etc.)

- **Separable utility:** Additive cost models, linear utility functions

- **Tree structure:** Hierarchical decision processes, causal models with tree structure

When a problem falls outside these cases, the hardness results apply, justifying heuristic approaches.

6 Why Over-Modeling Is Optimal

The complexity results of Sections 3 and 4 transform engineering practice from art to mathematics. This section proves that observed behaviors—configuration over-specification, absence of automated minimization tools, heuristic model selection—are not failures of discipline but *provably optimal responses* to computational constraints.

The conventional critique of over-modeling ("you should identify only the essential variables") is computationally naive. It asks engineers to solve coNP-complete problems. The rational response is to include everything and pay linear maintenance costs, rather than attempt exponential minimization costs.

6.1 Configuration Simplification is SUFFICIENCY-CHECK

Real engineering problems reduce directly to the decision problems studied in this paper.

Theorem 6.1 (Configuration Simplification Reduces to SUFFICIENCY-CHECK). *Given a software system with configuration parameters $P = \{p_1, \ldots, p_n\}$ and observed behaviors $B = \{b_1, \ldots, b_m\}$, the problem of determining whether parameter subset $I \subseteq P$ preserves all behaviors is equivalent to SUFFICIENCY-CHECK.*

Proof. Construct decision problem $\mathcal{D} = (A, X_1, \ldots, X_n, U)$ where:

- Actions $A = B$ (each behavior is an action)

- Coordinates $X_i =$ domain of parameter p_i

- State space $S = X_1 \times \cdots \times X_n$

- Utility $U(b, s) = 1$ if behavior b occurs under configuration s, else $U(b, s) = 0$

Then $\text{Opt}(s) = \{b \in B : b \text{ occurs under configuration } s\}$.
Coordinate set I is sufficient iff:

$$s_I = s'_I \implies \text{Opt}(s) = \text{Opt}(s')$$

This holds iff configurations agreeing on parameters in I exhibit identical behaviors.
Therefore, "does parameter subset I preserve all behaviors?" is exactly SUFFICIENCY-CHECK for the constructed decision problem. ∎

Remark 6.2. This reduction is *parsimonious*: every instance of configuration simplification corresponds bijectively to an instance of SUFFICIENCY-CHECK. The problems are not merely related—they are identical up to encoding.

6.2 Computational Rationality of Over-Modeling

We now prove that over-specification is the optimal engineering strategy given complexity constraints.

Theorem 6.3 (Rational Over-Modeling). *Consider an engineer specifying a system configuration with n parameters. Let:*

- $C_{over}(k)$ = *cost of maintaining k extra parameters beyond minimal*

- $C_{find}(n)$ = *cost of finding minimal sufficient parameter set*

- C_{under} = *expected cost of production failures from underspecification*

When SUFFICIENCY-CHECK is coNP-complete (Theorem 3.6):

1. *Worst-case finding cost is exponential: $C_{find}(n) = \Omega(2^n)$*

2. *Maintenance cost is linear: $C_{over}(k) = O(k)$*

3. *For sufficiently large n, exponential cost dominates linear cost*

Therefore, when n exceeds a threshold, over-modeling minimizes total expected cost:

$$C_{over}(k) < C_{find}(n) + C_{under}$$

Over-modeling is the economically optimal strategy under computational constraints.

Proof. By Theorem 3.6, SUFFICIENCY-CHECK is coNP-complete. Under standard complexity assumptions ($P \neq coNP$), no polynomial-time algorithm exists for checking sufficiency.

Finding the minimal sufficient set requires checking sufficiency of multiple candidate sets. Exhaustive search examines:

$$\sum_{i=0}^{n} \binom{n}{i} = 2^n \text{ candidate subsets}$$

Each check requires $\Omega(1)$ time (at minimum, reading the input). Therefore:

$$C_{\text{find}}(n) = \Omega(2^n)$$

Maintaining k extra parameters incurs:

- Documentation cost: $O(k)$ entries

- Testing cost: $O(k)$ test cases

- Migration cost: $O(k)$ parameters to update

Total maintenance cost is $C_{\text{over}}(k) = O(k)$.

For concrete threshold: when $n = 20$ parameters, exhaustive search requires $2^{20} \approx 10^6$ checks. Including $k = 5$ extra parameters costs $O(5)$ maintenance overhead but avoids 10^6 computational work.

Since 2^n grows faster than any polynomial in k or n, there exists n_0 such that for all $n > n_0$:

$$C_{\text{over}}(k) \ll C_{\text{find}}(n)$$

Adding underspecification risk C_{under} (production failures from missing parameters), which can be arbitrarily large, makes over-specification strictly dominant. ∎

Corollary 6.4 (Impossibility of Automated Configuration Minimization). *There exists no polynomial-time algorithm that:*

1. *Takes an arbitrary configuration file with n parameters*

2. *Identifies the minimal sufficient parameter subset*

3. *Guarantees correctness (no false negatives)*

Proof. Such an algorithm would solve MINIMUM-SUFFICIENT-SET in polynomial time, contradicting Theorem 3.7 (assuming $P \neq coNP$). ∎

Remark 6.5. Corollary 6.4 explains the observed absence of "config cleanup" tools in software engineering practice. Engineers who include extra parameters are not exhibiting poor discipline—they are adapting optimally to computational impossibility. The problem is not lack of tooling effort; it is mathematical intractability.

6.3 Connection to Observed Practice

These theorems provide mathematical grounding for three widespread engineering behaviors:

1. Configuration files grow over time. Removing parameters requires solving coNP-complete problems. Engineers rationally choose linear maintenance cost over exponential minimization cost.

2. Heuristic model selection dominates. ML practitioners use AIC, BIC, cross-validation instead of optimal feature selection because optimal selection is intractable (Theorem 6.3).

3. "Include everything" is a legitimate strategy. When determining relevance costs $\Omega(2^n)$, including all n parameters costs $O(n)$. For large n, this is the rational choice.

These are not workarounds or approximations. They are *optimal responses* to computational constraints. The complexity results transform engineering practice from art to mathematics: over-modeling is not a failure—it is the provably correct strategy.

7 Implications for Software Architecture

The complexity results have direct implications for software engineering practice.

7.1 Why Over-Specification Is Rational

Software architects routinely specify more configuration parameters than strictly necessary. Our results show this is computationally rational:

Corollary 7.1 (Rational Over-Specification). *Given a software system with n configuration parameters, checking whether a proposed subset suffices is coNP-complete. Finding the minimum such set is also coNP-complete.*

This explains why configuration files grow over time: removing "unnecessary" parameters requires solving a hard problem.

7.2 Connection to Leverage Theory

Paper 3 introduced leverage as the ratio of impact to effort. The decision quotient provides a complementary measure:

Definition 7.2 (Architectural Decision Quotient)**.** For a software system with configuration space S and behavior space B:

$$\text{ADQ}(I) = \frac{|\{b \in B : b \text{ achievable with some } s \text{ where } s_I \text{ fixed}\}|}{|B|}$$

High ADQ means the configuration subset I leaves many behaviors achievable—it doesn't constrain the system much. Low ADQ means I strongly constrains behavior.

Proposition 7.3 (Leverage-ADQ Duality)**.** *High-leverage architectural decisions correspond to low-ADQ configuration subsets: they strongly constrain system behavior with minimal specification.*

7.3 Practical Recommendations

Based on our theoretical results:

1. **Accept over-modeling:** Don't penalize engineers for including "extra" parameters. The alternative (minimal modeling) is computationally hard.

2. **Use bounded scenarios:** When the scenario space is small (Proposition 2.10), minimal modeling becomes tractable.

3. **Exploit structure:** Tree-structured dependencies, bounded alternatives, and separable utilities admit efficient algorithms.

4. **Invest in heuristics:** For general problems, develop domain-specific heuristics rather than seeking optimal solutions.

7.4 Hardness Distribution: Right Place vs Wrong Place

A fundamental principle emerges from the complexity results: problem hardness is conserved but can be *distributed* across a system in qualitatively different ways.

Definition 7.4 (Hardness Distribution)**.** Let P be a problem with intrinsic hardness $H(P)$ (measured in computational steps, implementation effort, or error probability). A *solution architecture* S partitions this hardness into:

- $H_{\text{central}}(S)$: hardness paid once, at design time or in a shared component

- $H_{\text{distributed}}(S)$: hardness paid per use site

For n use sites, total realized hardness is:

$$H_{\text{total}}(S) = H_{\text{central}}(S) + n \cdot H_{\text{distributed}}(S)$$

Theorem 7.5 (Hardness Conservation)**.** *For any problem P with intrinsic hardness $H(P)$, any solution S satisfies:*

$$H_{central}(S) + H_{distributed}(S) \geq H(P)$$

Hardness cannot be eliminated, only redistributed.

Proof. By definition of intrinsic hardness: any correct solution must perform at least $H(P)$ units of work (computational, cognitive, or error-handling). This work is either centralized or distributed. ∎ ∎

Definition 7.6 (Hardness Efficiency)**.** The *hardness efficiency* of solution S with n use sites is:

$$\eta(S, n) = \frac{H_{\text{central}}(S)}{H_{\text{central}}(S) + n \cdot H_{\text{distributed}}(S)}$$

High η indicates centralized hardness (paid once); low η indicates distributed hardness (paid repeatedly).

Theorem 7.7 (Centralization Dominance)**.** *For $n > 1$ use sites, solutions with higher $H_{central}$ and lower $H_{distributed}$ yield:*

1. *Lower total realized hardness: $H_{total}(S_1) < H_{total}(S_2)$ when $H_{distributed}(S_1) < H_{distributed}(S_2)$*

2. *Fewer error sites: errors in centralized components affect 1 location; errors in distributed components affect n locations*

3. *Higher leverage (Paper 3): one unit of central effort affects n sites*

Proof. (1) follows from the total hardness formula. (2) follows from error site counting. (3) follows from Paper 3's leverage definition $L = \Delta\text{Effect}/\Delta\text{Effort}$. ∎ ∎

Corollary 7.8 (Right Hardness vs Wrong Hardness)**.** *A solution exhibits* hardness in the right place *when:*

- *Hardness is centralized (high $H_{central}$, low $H_{distributed}$)*

- *Hardness is paid at design/compile time rather than runtime*

- *Hardness is enforced by tooling (type checker, compiler) rather than convention*

A solution exhibits hardness in the wrong place *when:*

- *Hardness is distributed (low $H_{central}$, high $H_{distributed}$)*

- *Hardness is paid repeatedly at each use site*

- *Hardness relies on human discipline rather than mechanical enforcement*

Example: Type System Instantiation. Consider a capability C (e.g., provenance tracking) that requires hardness $H(C)$:

Approach	H_{central}	$H_{\text{distributed}}$
Native type system support	High (learning cost)	Low (type checker enforces)
Manual implementation	Low (no new concepts)	High (reimplement per site)

For n use sites, manual implementation costs $n \cdot H_{\text{distributed}}$, growing without bound. Native support costs H_{central} once, amortized across all uses. The "simpler" approach (manual) is only simpler at $n = 1$; for $n > H_{\text{central}}/H_{\text{distributed}}$, native support dominates.

Remark 7.9 (Connection to Decision Quotient). The decision quotient (Section 2) measures which coordinates are decision-relevant. Hardness distribution measures where the cost of *handling* those coordinates is paid. A high-axis system makes relevance explicit (central hardness); a low-axis system requires users to track relevance themselves (distributed hardness).

The next section develops the major practical consequence of this framework: the Simplicity Tax Theorem.

8 The Simplicity Tax Theorem

The complexity results of Sections 3–5 establish that identifying decision-relevant dimensions is coNP-complete. This section develops the practical consequence: what happens when engineers *ignore* this hardness and attempt to use "simple" tools for complex problems.

The answer is the *Simplicity Tax*: a per-site cost that cannot be avoided, only redistributed. This result overturns the common intuition that "simpler is always better" and establishes a foundational principle: **No Free Simplicity**.

8.1 The Conservation Law: No Free Simplicity

Definition 8.1 (Problem and Tool). A *problem P* has a set of *required axes $R(P)$*—the dimensions of variation that must be represented. A *tool T* has a set of *native axes $A(T)$*—what it can represent directly.

This terminology is grounded in Papers 1–2: "axes" correspond to Paper 1's axis framework (`requiredAxesOf`) and Paper 2's degrees of freedom.

Definition 8.2 (Expressive Gap and Simplicity Tax). The *expressive gap* between tool T and problem P is:
$$\mathrm{Gap}(T, P) = R(P) \setminus A(T)$$
The *simplicity tax* is $|\mathrm{Gap}(T, P)|$: the number of axes the tool cannot handle natively. This tax is paid at *every use site*.

Definition 8.3 (Complete vs. Incomplete Tools). Tool T is *complete* for problem P if $R(P) \subseteq A(T)$. Otherwise T is *incomplete* for P.

Theorem 8.4 (Simplicity Tax Conservation). *For any problem P with required axes $R(P)$ and any tool T:*
$$|Gap(T, P)| + |R(P) \cap A(T)| = |R(P)|$$
The required axes are partitioned into "covered natively" and "tax." You cannot reduce the total—only shift where it is paid.

Proof. Set partition: $R(P) = (R(P) \cap A(T)) \cup (R(P) \setminus A(T))$. The sets are disjoint. Cardinality follows. ∎ ∎

This is analogous to conservation laws in physics: energy is conserved, only transformed. Complexity is conserved, only distributed.

Theorem 8.5 (Complete Tools Pay No Tax). *If T is complete for P, then $SimplicityTax(T, P) = 0$.*

Theorem 8.6 (Incomplete Tools Pay Positive Tax). *If T is incomplete for P, then $SimplicityTax(T, P) > 0$.*

Theorem 8.7 (Simplicity Tax Grows Linearly). *For n use sites with an incomplete tool:*
$$TotalExternalWork(T, P, n) = n \times SimplicityTax(T, P)$$

Total work grows linearly. There is no economy of scale for distributed complexity.

Theorem 8.8 (Complete Dominates Incomplete). *For any $n > 0$, a complete tool has strictly less total cost than an incomplete tool:*
$$TotalExternalWork(T_{complete}, P, n) < TotalExternalWork(T_{incomplete}, P, n)$$

Proof. Complete: $0 \times n = 0$. Incomplete: $k \times n$ for $k \geq 1$. For $n > 0$: $0 < kn$. ∎ ∎

8.2 The Simplicity Preference Fallacy

Definition 8.9 (Simplicity Preference Fallacy)**.** The *simplicity preference fallacy* is the cognitive error of preferring low H_{central} (learning cost) without accounting for $H_{\text{distributed}}$ (per-site cost).

This fallacy manifests as:

- "I prefer simple tools" (without asking: simple relative to what problem?)

- "YAGNI" applied to infrastructure (ignoring amortization across use sites)

- "Just write straightforward code" (ignoring that n sites pay the tax)

- "Abstractions are overhead" (treating central cost as total cost)

Theorem 8.10 (The Fallacy Theorem)**.** *Let T_{simple} be incomplete for problem P and $T_{complex}$ be complete. For any $n > 0$:*

$$TotalExternalWork(T_{complex}, P, n) < TotalExternalWork(T_{simple}, P, n)$$

The "simpler" tool creates more total work, not less.

The fallacy persists because $H_{\text{distributed}}$ is invisible: it is paid by users, at runtime, across time, in maintenance. H_{central} is visible: it is paid by the designer, upfront, once. Humans overweight visible costs.

Theorem 8.11 (Amortization Threshold)**.** *There exists a threshold n^* such that for all $n > n^*$, the total cost of the "complex" tool (including learning) is strictly less than the "simple" tool:*

$$n^* = \frac{H_{central}(T_{complex})}{SimplicityTax(T_{simple}, P)}$$

Beyond n^* uses, the complex tool is cheaper even accounting for learning cost.

Remark 8.12 (On the Learning Cost Model). This theorem models learning cost as H_{central}, a scalar. A more precise formalization would treat learning cost as the rank of a *concept matroid*—the prerequisite concepts required to master the tool (see Conclusion, Future Work). Paper 1 established that type axes form matroids; concept axes may admit similar structure. Critically, the matroid property ensures that *different minimal learning paths have equal cardinality*, making the scalar well-defined despite multiple valid trajectories. The qualitative result (amortization threshold exists) is robust to the learning cost model; the quantitative threshold depends on its precise formalization.

8.3 Cross-Domain Examples

The Simplicity Tax applies universally. Consider these domains:

Domain	"Simple" Choice	"Complex" Choice	Tax per Site
Type Systems	Dynamic typing	Static typing	Runtime type errors
Python	Manual patterns	Metaclasses/descriptors	Boilerplate code
Data Validation	Ad-hoc checks	Schema/ORM	Validation logic
Configuration	Hardcoded values	Config management	Change propagation
APIs	Stringly-typed	Rich type models	Parse/validate code

In each case, the "simple" choice has lower learning cost (H_{central}) but higher per-site cost ($H_{\text{distributed}}$). For n use sites, the simple choice costs $n \times$ tax.

Example: Python Metaclasses. Python's community often resists metaclasses as "too complex." But consider a problem requiring automatic subclass registration, attribute validation, and interface enforcement—three axes of variation.

Approach	Native Axes	Tax/Class	Total for 50 classes
Metaclass	{registration, validation, interface}	0	0
Manual decorators	{registration}	2	100
Fully manual	∅	3	150

The "simplest" approach (fully manual) creates the most work. The community's resistance to metaclasses is the Simplicity Preference Fallacy in action.

Example: Static vs. Dynamic Typing. Dynamic typing has lower learning cost. But type errors are a per-site tax: each call site that could receive a wrong type is an error site. For n call sites:

- Static typing: type checker verifies once, 0 runtime type errors

- Dynamic typing: n potential runtime type errors, each requiring defensive code or debugging

The "simplicity" of dynamic typing distributes type-checking to every call site.

8.4 Unification with Papers 1–3

The Simplicity Tax Theorem unifies results across the pentalogy:

Paper 1 (Typing Disciplines). Fixed-axis type systems are incomplete for domains requiring additional axes. The Simplicity Tax quantifies the cost: $|\text{requiredAxes}(D) \setminus \text{fixedAxes}|$ per use site. Parameterized type systems are complete (zero tax).

Paper 2 (SSOT). Non-SSOT architectures distribute specification across n locations. Each location is a potential error site. SSOT centralizes specification: $H_{\text{distributed}} = 0$.

Paper 3 (Leverage). High-leverage solutions have high H_{central} and low $H_{\text{distributed}}$. Leverage $=$ impact/effort $= n/H_{\text{central}}$ when $H_{\text{distributed}} = 0$. Low-leverage solutions pay per-site.

Paper 4 (This Paper). Identifying which axes matter is coNP-complete. If you guess wrong and use an incomplete tool, you pay the Simplicity Tax. The tax is the cost of the coNP-hard problem you failed to solve.

Theorem 8.13 (Unified Dominance). *Across Papers 1–4, solutions with higher $H_{central}$ and zero $H_{distributed}$ strictly dominate solutions with lower $H_{central}$ and positive $H_{distributed}$, for $n > n^*$.*

These are not four separate claims. They are four views of a single phenomenon: the conservation and distribution of intrinsic problem complexity.

8.5 Formal Competence

Definition 8.14 (Formal Competence). An engineer is *formally competent* with respect to complexity distribution if they correctly account for both H_{central} and $H_{\text{distributed}}$ when evaluating tools.

The Competence Test:

1. Identify intrinsic problem complexity: $|R(P)|$

2. Identify tool's native axes: $|A(T)|$

3. Compute the gap: $|R(P) \setminus A(T)|$

4. Compute total cost: $H_{\text{central}}(T) + n \times |R(P) \setminus A(T)|$

5. Compare tools by total cost, not H_{central} alone

Failing step 4—evaluating tools by learning cost alone—is formal incompetence.

Remark 8.15 (The Zen of Python, Correctly Read). Python's Zen states: "Simple is better than complex. Complex is better than complicated." This is often misread as endorsing simplicity unconditionally. The correct reading:

- **Simple**: Low intrinsic complexity (both H_{central} and $H_{\text{distributed}}$ low)

- **Complex**: High intrinsic complexity, *structured* (high H_{central}, low $H_{\text{distributed}}$)

- **Complicated**: High intrinsic complexity, *tangled* (low H_{central}, high $H_{\text{distributed}}$)

The Zen says: when the problem has intrinsic complexity, *complex* (centralized) beats *complicated* (distributed). The community often conflates complex with complicated.

8.6   Lean 4 Formalization

All theorems in this section are machine-checked in `DecisionQuotient/HardnessDistribution.lean`:

Theorem	Lean Name
Simplicity Tax Conservation	`simplicityTax_conservation`
Complete Tools Pay No Tax	`complete_tool_no_tax`
Incomplete Tools Pay Positive Tax	`incomplete_tool_positive_tax`
Tax Grows Linearly	`simplicityTax_grows`
Complete Dominates Incomplete	`complete_dominates_incomplete`
The Fallacy Theorem	`simplicity_preference_fallacy`
Amortization Threshold	`amortization_threshold`
Dominance Transitivity	`dominates_trans`
Tax Antitone w.r.t. Expressiveness	`simplicityTax_antitone`

The formalization uses `Finset` $\mathbb{N}$ for axes, making the simplicity tax a computable natural number. The `Tool` type forms a lattice under the expressiveness ordering, with tax antitone (more expressive $\Rightarrow$ lower tax).

All proofs compile with zero `sorry` placeholders.

9   Related Work

9.1   Computational Decision Theory

The complexity of decision-making has been studied extensively. Papadimitriou [19] established foundational results on the complexity of game-theoretic solution concepts. Our work extends this to the meta-question of identifying relevant information. For a modern treatment of complexity classes, see Arora and Barak [3].

9.2 Feature Selection Complexity

In machine learning, feature selection asks which input features are relevant for prediction. Blum and Langley [4] survey the field, noting hardness in general settings. Amaldi and Kann [2] proved that finding minimum feature sets for linear classifiers is NP-hard, and established inapproximability bounds: no polynomial-time algorithm can approximate the minimum feature set within factor $2^{\log^{1-\epsilon} n}$ unless NP $\subseteq$ DTIME($n^{\text{polylog } n}$).

Our results extend this line: the decision-theoretic analog (SUFFICIENCY-CHECK) is coNP-complete, and MINIMUM-SUFFICIENT-SET inherits this hardness. The key insight is that sufficiency checking is "dual" to feature selection—rather than asking which features predict a label, we ask which coordinates determine optimal action. The coNP (rather than NP) classification reflects this duality: insufficiency has short certificates (counterexample state pairs), while sufficiency requires universal verification.

9.3 Sufficient Statistics

Fisher [9] introduced sufficient statistics: a statistic $T(X)$ is *sufficient* for parameter θ if the conditional distribution of X given $T(X)$ does not depend on θ. Lehmann and Scheffé [16] characterized minimal sufficient statistics and their uniqueness properties.

Our coordinate sufficiency is the decision-theoretic analog: a coordinate set I is sufficient if knowing s_I determines optimal action, regardless of the remaining coordinates. The parallel is precise:

- **Statistics:** T is sufficient $\iff P(X|T(X), \theta) = P(X|T(X))$

- **Decisions:** I is sufficient $\iff \text{Opt}(s) = \text{Opt}(s')$ whenever $s_I = s'_I$

Fisher's factorization theorem provides a characterization; our Theorem 3.7 shows that *finding* minimal sufficient statistics (in the decision-theoretic sense) is computationally hard.

9.4 Causal Inference and Adjustment Sets

Pearl [21] and Spirtes et al. [28] developed frameworks for identifying causal effects from observational data. A central question is: which variables must be adjusted for to identify a causal effect? The *adjustment criterion* and *back-door criterion* characterize sufficient adjustment sets.

Our sufficiency problem is analogous: which coordinates must be observed to determine optimal action? The complexity results suggest that optimal adjustment set selection may also be intractable—a conjecture supported by recent work on the complexity of causal discovery [5].

The connection runs deeper: Shpitser and Pearl [26] showed that identifying causal effects is NP-hard in general graphs. Our coNP-completeness result for SUFFICIENCY-CHECK is the decision-theoretic counterpart.

9.5 Minimum Description Length and Kolmogorov Complexity

The Minimum Description Length (MDL) principle [23, 10] formalizes model selection as compression: the best model minimizes description length of data plus model. Kolmogorov complexity [17] provides the theoretical foundation—the shortest program that generates the data.

Our decision quotient connects to this perspective: a coordinate set I is sufficient if it compresses the decision problem without loss—knowing s_I is as good as knowing s for decision purposes. The minimal sufficient set is the MDL-optimal compression of the decision problem.

The complexity results explain why MDL-based model selection uses heuristics: finding the true minimum description length is uncomputable (Kolmogorov complexity) or intractable (MDL approximations). Our results show the decision-theoretic analog is coNP-complete—intractable but decidable.

9.6 Value of Information

The value of information (VOI) framework [12] quantifies how much a decision-maker should pay for information. Our work addresses a different question: not the *value* of information, but the *complexity* of identifying which information has value.

Interestingly, VOI is typically polynomial to compute given the decision problem structure, while identifying which information *to value* (our problem) is coNP-complete. This separation explains why VOI is practical while optimal sensor placement remains heuristic.

9.7 Sensitivity Analysis

Sensitivity analysis asks how outputs change with inputs. Local sensitivity (derivatives) is polynomial; global sensitivity (Sobol indices [27]) requires sampling. Identifying which inputs *matter* for decision-making is our sufficiency problem—which we show is coNP-complete.

This explains why practitioners use sampling-based sensitivity analysis rather than exact methods: exact identification of decision-relevant inputs is intractable. The dichotomy theorem (Theorem 4.1) characterizes when sensitivity analysis becomes tractable (logarithmic relevant inputs) versus intractable (linear relevant inputs).

9.8 Model Selection

Statistical model selection (AIC [1], BIC [25], cross-validation [30]) provides practical heuristics for choosing among models. Our results provide theoretical justification: optimal model selection is intractable, so heuristics are necessary.

The Simplicity Tax Theorem (Section 8) adds a warning: model selection heuristics that favor "simpler" models may incur hidden costs when the true model is complex. The simplicity preference fallacy—choosing low-parameter models without accounting for per-site costs—is the decision-theoretic formalization of overfitting-by-underfitting.

10 Conclusion

Methodology and Disclosure

Role of LLMs in this work. This paper was developed through human-AI collaboration. The author provided the core intuitions—the connection between decision-relevance and computational complexity, the conjecture that SUFFICIENCY-CHECK is coNP-complete, and the insight that the Σ_2^P structure collapses for MINIMUM-SUFFICIENT-SET. Large language models (Claude, GPT-4) served as implementation partners for proof drafting, Lean formalization, and LaTeX generation.

The Lean 4 proofs were iteratively refined: the author specified what should be proved, the LLM proposed proof strategies, and the Lean compiler served as the arbiter of correctness. The complexity-theoretic reductions required careful human oversight to ensure the polynomial bounds were correctly established.

What the author contributed: The problem formulations (SUFFICIENCY-CHECK, MINIMUM-SUFFICIENT-SET, ANCHOR-SUFFICIENCY), the hardness conjectures, the tractability conditions, and the connection to over-modeling in engineering practice.

What LLMs contributed: LaTeX drafting, Lean tactic exploration, reduction construction assistance, and prose refinement.

The proofs are machine-checked; their validity is independent of generation method. We disclose this methodology in the interest of academic transparency.

We have established that identifying decision-relevant information is computationally hard:

- Checking whether a coordinate set is sufficient is coNP-complete

- Finding the minimum sufficient set is coNP-complete (the Σ_2^P structure collapses)

- Anchor sufficiency (fixed-coordinate subcube) is Σ_2^P-complete

- A complexity dichotomy separates easy (logarithmic) from hard (linear) cases

- Tractable subcases exist for bounded actions, separable utilities, and tree structures

These results establish a fundamental principle of rational decision-making under uncertainty:

Determining what you need to know is harder than knowing everything.

This is not a metaphor or heuristic observation. It is a mathematical theorem with universal scope. Any agent facing structured uncertainty—whether a climate scientist, financial analyst, software engineer, or artificial intelligence—faces the same computational constraint. The ubiquity of over-modeling across domains is not coincidence, laziness, or poor discipline. It is the provably optimal response to intractability.

The principle has immediate normative force: stop criticizing engineers for including "irrelevant" parameters. Stop demanding minimal models. Stop building tools that promise to identify "what really matters." These aspirations conflict with computational reality. The dichotomy theorem (Theorem 4.1) characterizes exactly when tractability holds; outside those boundaries, over-modeling is not a failure mode—it is the only rational strategy.

All proofs are machine-checked in Lean 4, ensuring correctness of the core mathematical claims including the reduction mappings and equivalence theorems. Complexity classifications follow from standard complexity-theoretic results (TAUTOLOGY is coNP-complete, $\exists\forall$-SAT is Σ_2^P-complete) under the encoding model described in Section 3.

Why These Results Are Final

The theorems proven here are *ceiling results*—no stronger claims are possible within their respective frameworks:

1. **Exact complexity characterization (not just lower bounds).** We prove SUFFICIENCY-CHECK is coNP-complete (Theorem 3.6). This is *both* a lower bound (coNP-hard) and an upper bound (in coNP). The complexity class is exact. Additional lower or upper bounds would be redundant.

2. **Universal impossibility ($\forall$), not probabilistic prevalence ($\mu = 1$).** Theorems quantify over *all* decision problems satisfying the structural constraints, not measure-1 subsets. Measure-theoretic claims like "hard instances are prevalent" would *weaken* the results from "always hard (unless P = coNP)" to "almost always hard."

3. **Constructive reductions, not existence proofs.** Theorem 3.6 provides an explicit polynomial-time reduction from TAUTOLOGY to SUFFICIENCY-CHECK. This is stronger than proving hardness via non-constructive arguments (e.g., diagonalization). The reduction is machine-checked and executable.

4. **Dichotomy is complete (Theorem 4.1).** The complexity separates into exactly two cases: polynomial (when minimal sufficient set has size $O(\log |S|)$) or exponential (when size is $\Omega(n)$). Under standard assumptions (P $\neq$ coNP), there are no intermediate cases. The dichotomy is exhaustive.

5. **Tractable cases are maximal (Section 5).** The tractability conditions (bounded actions, separable utilities, tree structure) are shown to be *tight*—relaxing any condition yields coNP-hardness. These are the boundary cases, not a subset of tractable instances.

What would NOT strengthen the results:

- **Additional complexity classes:** SUFFICIENCY-CHECK is coNP-complete. Proving it is also NP-hard, PSPACE-hard, or #P-hard would add no information (these follow from coNP-completeness under standard reductions).

- **Average-case hardness:** We prove worst-case hardness. Average-case results would be *weaker* (average $\leq$ worst) and would require distributional assumptions not present in the problem definition.

- **#P-hardness of counting:** When the decision problem is asking "does there exist?" (existential) or "are all?" (universal), the corresponding counting problem is trivially at least as hard. Proving #P-hardness separately would be redundant unless we change the problem to count something else.

- **Approximation hardness beyond inapproximability:** The coNP-completeness of MINIMUM-SUFFICIENT-SET (Theorem 3.7) implies no polynomial-time algorithm can approximate the minimal sufficient set size within any constant factor (unless P = coNP). This is maximal inapproximability—the problem admits no non-trivial approximation.

These results close the complexity landscape for coordinate sufficiency. Within classical complexity theory, the characterization is complete.

The Simplicity Tax: A Major Practical Consequence

A widespread belief holds that "simpler is better"—that preferring simple tools and minimal models is a mark of sophistication. This paper proves that belief is context-dependent and often wrong.

The *Simplicity Tax Theorem* (Section 8) establishes: when a problem requires k axes of variation and a tool natively supports only $j < k$ of them, the remaining $k-j$ axes must be handled externally at *every use site*. For n use sites, the "simpler" tool creates $(k - j) \times n$ units of external work. A tool matched to the problem's complexity creates zero external work.

True sophistication is matching tool complexity to problem complexity.

Preferring "simple" tools for complex problems is not wisdom—it is a failure to account for distributed costs. The simplicity tax is paid invisibly, at every use site, by every user, forever. The sophisticated engineer asks not "which tool is simpler?" but "which tool matches my problem's intrinsic complexity?"

This result is machine-checked in Lean 4 (`HardnessDistribution.lean`). The formalization proves conservation (you can't eliminate the tax, only redistribute it), dominance (complete tools always beat incomplete tools), and the amortization threshold (beyond which the "complex" tool is strictly cheaper).

The Foundational Contribution

This paper proves a universal constraint on optimization under uncertainty. The constraint is:

- **Mathematical**, not empirical—it follows from the structure of computation

- **Universal**, not domain-specific—it applies to any decision problem with coordinate structure

- **Permanent**, not provisional—no algorithmic breakthrough can circumvent `coNP`-completeness (unless P = coNP)

The result explains phenomena across disciplines: why feature selection uses heuristics, why configuration files grow, why sensitivity analysis is approximate, why model selection is art rather than science. These are not separate problems with separate explanations. They are manifestations of a single computational constraint, now formally characterized.

The Simplicity Tax Theorem adds a practical corollary: the universal response to intractability (over-modeling) is not just rational—attempting to avoid it by using "simpler" tools is actively counterproductive. Simplicity that mismatches problem complexity creates more work, not less.

Open questions remain (fixed-parameter tractability, quantum complexity, average-case behavior under natural distributions), but the foundational question—*is identifying relevance fundamentally hard?*—is answered: yes.

Future Work: Learning Cost as Concept Matroids

The Simplicity Tax Theorem uses a simplified model of learning cost (H_{central}). A more precise formalization would treat learning cost as *concept axes*—the prerequisite concepts required to use a tool effectively.

Paper 1 established that type axes form a matroid structure: minimal complete axis sets have equal cardinality, and the `type()` operation witnesses axis membership. An analogous structure may govern learning:

- Each prerequisite concept (classes, inheritance, descriptors, MRO, etc.) is an axis

- **Different minimal learning paths exist with equal cardinality** (the matroid property)

- Learning cost = rank of the concept matroid (size of any minimal generating set)

- The primitive witnessing concept mastery = **Minimal Working Example (MWE)**

The witness deserves elaboration. Paper 1 used `type(x)` as the primitive witnessing type-axis membership. For concept axes, the analogue is the *Minimal Working Example*—the shortest program demonstrating correct usage of a capability. This is precisely the Kolmogorov complexity of competence, measured in AST nodes rather than bits:

$$\text{LearningCost}(T) = \min\{|\text{AST}(p)| : p \text{ correctly uses capability } T\}$$

For `type()`: the MWE is `type(x)` ($\sim$1 node). For metaclasses: the MWE requires class definition, inheritance from `type`, `__new__` override, and `super()` delegation ($\sim$15 nodes). The ratio of AST node counts is the ratio of learning costs.

The matroid structure is non-trivial: to learn metaclasses, one might traverse concepts via inheritance and `__new__`, or via descriptors and `__call__`, or via class decorators. These are *different* minimal paths, but the exchange property guarantees they have the same length. This is precisely why learning cost can be a well-defined scalar (the rank) despite multiple valid learning trajectories.

This yields a three-level hierarchy:

1. **Problem axes**: what the domain requires

2. **Tool axes**: what the tool natively handles

3. **Concept axes**: what must be learned to access tool axes

Complete mastery requires: concept axes $\supseteq$ tool axes $\supseteq$ problem axes.

The amortization threshold (Theorem 8.11) would then have a precise characterization: $n^* = \text{rank}(\text{ConceptMatroid}(T))/\text{SimplicityTax}(T, P)$.

Full development of concept matroids is deferred to future work. The key conjecture: *learning cost, like type structure, admits matroid formalization, making the Simplicity Tax quantitatively precise rather than merely qualitatively correct.*

A   Lean 4 Proof Listings

The complete Lean 4 formalization is available at:

https://doi.org/10.5281/zenodo.18140966

A.1   On the Nature of Foundational Proofs

The Lean proofs are straightforward applications of definitions and standard complexity-theoretic constructions. Foundational work produces insight through formalization.

Definitional vs. derivational proofs. The core theorems establish definitional properties and reduction constructions. For example, the polynomial reduction composition theorem (Theorem A.1) proves that composing two polynomial-time reductions yields a polynomial-time reduction. The proof follows from the definition of polynomial time: composing two polynomials yields a polynomial.

Precedent in complexity theory. This pattern appears throughout foundational complexity theory:

- **Cook-Levin Theorem (1971):** SAT is NP-complete. The proof constructs a reduction from an arbitrary NP problem to SAT. The construction itself is straightforward (encode Turing machine computation as boolean formula), but the *insight* is recognizing that SAT captures all of NP.

- **Ladner's Theorem (1975):** If P $\neq$ NP, then NP-intermediate problems exist. The proof is a diagonal construction—conceptually simple once the right framework is identified.

- **Toda's Theorem (1991):** The polynomial hierarchy is contained in P$^{\#}$P. The proof uses counting arguments that are elegant but not technically complex. The profundity is in the *connection* between counting and the hierarchy.

Why simplicity indicates strength. A definitional theorem derived from precise formalization is *stronger* than an informal argument. When we prove that sufficiency checking is coNP-complete (Theorem 3.6), we are not saying "we tried many algorithms and they all failed." We are saying something universal: *any* algorithm solving sufficiency checking can solve TAUTOLOGY, and vice versa. The proof is a reduction construction that follows from the problem definitions.

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

1. **Precision forcing.** Formalizing "coordinate sufficiency" in Lean requires stating exactly what it means for a coordinate subset to contain all decision-relevant information. This precision eliminates ambiguity about edge cases (what if projections differ only on irrelevant coordinates?).

2. **Reduction correctness.** The TAUTOLOGY reduction (Section 3) is machine-checked to preserve the decision structure. Informal reductions can have subtle bugs; Lean verification guarantees the mapping is correct.

3. **Complexity dichotomy.** Theorem 4.1 proves that problem instances are either tractable (P) or intractable (coNP-complete), with no intermediate cases under standard assumptions. This emerges from the formalization of constraint structure, not from case enumeration.

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 (extended with Mathlib for basic combinatorics and complexity definitions). Zero `sorry` placeholders means zero unproven claims. The 3,400+ lines establish a verified chain from basic definitions (decision problems, coordinate spaces, polynomial reductions) to the final theorems (hardness results, dichotomy, tractable cases). Reviewers need not trust our informal explanations—they can run `lake build` and verify the proofs themselves.

Comparison to informal complexity arguments. Prior work on model selection complexity (Chickering et al. [5], Teyssier & Koller [31]) presents compelling informal arguments but lacks machine-checked proofs. Our contribution is not new *wisdom*—the insight that model selection is hard is old. Our contribution is *formalization*: making "coordinate sufficiency" precise enough to mechanize, constructing verified reductions, and proving the complexity results hold for the formalized definitions.

This follows the tradition of verified complexity theory: just as Nipkow & Klein [18] formalized automata theory and Cook [7] formalized NP-completeness in proof assistants, we formalize decision-theoretic complexity. The proofs are simple because the formalization makes the structure clear. Simple proofs from precise definitions are the goal, not a limitation.

A.2 Module Structure

The formalization consists of 33 files organized as follows:

- `Basic.lean` – Core definitions (DecisionProblem, CoordinateSet, Projection)

- `AlgorithmComplexity.lean` – Complexity definitions (polynomial time, reductions)

- `PolynomialReduction.lean` – Polynomial reduction composition (Theorem A.1)

- `Reduction.lean` – TAUTOLOGY reduction for sufficiency checking

- `Hardness/` – Counting complexity and approximation barriers

- `Tractability/` – Bounded actions, separable utilities, tree structure, FPT

- `Economics/` – Value of information and elicitation connections

- `Dichotomy.lean` and `ComplexityMain.lean` – Summary results

- `HardnessDistribution.lean` – Simplicity Tax Theorem (Section 8)

A.3 Key Theorems

Theorem A.1 (Polynomial Composition, Lean). *Polynomial-time reductions compose to polynomial-time reductions.*

```
theorem PolyReduction.comp_exists
    (f : PolyReduction A B) (g : PolyReduction B C) :
    exists h : PolyReduction A C,
      forall a, h.reduce a = g.reduce (f.reduce a)
```

Theorem A.2 (Simplicity Tax Conservation, Lean). *The simplicity tax plus covered axes equals required axes (partition).*

```
theorem simplicityTax_conservation :
    simplicityTax P T + (P.requiredAxes inter T.nativeAxes).card
      = P.requiredAxes.card
```

Theorem A.3 (Simplicity Preference Fallacy, Lean). *Incomplete tools always cost more than complete tools for $n > 0$ use sites.*

```
theorem simplicity_preference_fallacy (T_simple T_complex : Tool)
    (h_simple_incomplete : isIncomplete P T_simple)
    (h_complex_complete : isComplete P T_complex)
    (n : Nat) (hn : n > 0) :
    totalExternalWork P T_complex n < totalExternalWork P T_simple n
```

A.4 Verification Status

- Total lines: $\sim$5,000

- Theorems: 200+

- Files: 33

- Status: All proofs compile with no `sorry`

B Preemptive Rebuttals

We address anticipated objections to the main results.

B.1 Objection 1: "coNP-completeness doesn't mean intractable"

Objection: "coNP-complete problems might have good heuristics or approximations. The hardness result doesn't preclude practical solutions."

Response: This objection actually *strengthens* our thesis. The point is not that practitioners cannot find useful approximations—they clearly do (feature selection heuristics in ML, sensitivity analysis in economics, configuration defaults in software). The point is that *optimal* dimension selection is provably hard.

The prevalence of heuristics across domains is itself evidence of the computational barrier. If optimal selection were tractable, we would see optimal algorithms, not heuristics. The universal adoption of "include more than necessary" strategies is the rational response to coNP-completeness.

B.2 Objection 2: "Real problems don't have coordinate structure"

Objection: "Real decision problems are messier than your clean product-space model. The coordinate structure assumption is too restrictive."

Response: The assumption is weaker than it appears. Any finite state space can be encoded with binary coordinates; our hardness results apply to this encoding. More structured representations make the problem *easier*, not harder—so hardness for structured problems implies hardness for unstructured ones.

The coordinate structure abstracts common patterns: independent sensors, orthogonal configuration parameters, factored state spaces. These are ubiquitous in practice precisely because they enable tractable reasoning in special cases (Theorem 5.1).

B.3 Objection 3: "The SAT reduction is artificial"

Objection: "The reduction from SAT/TAUT is an artifact of complexity theory. Real decision problems don't encode Boolean formulas."

Response: All coNP-completeness proofs use reductions. The reduction demonstrates that TAUT instances can be encoded as sufficiency-checking problems while preserving computational structure. This is standard methodology [6, 14].

The claim is not that practitioners encounter SAT problems in disguise, but that sufficiency checking is *at least as hard as* TAUT. If sufficiency checking were tractable, we could solve TAUT in polynomial time, contradicting the widely-believed $\P \neq \NP$ conjecture.

The reduction is a proof technique, not a claim about problem origins.

B.4 Objection 4: "Tractable subcases are too restrictive"

Objection: "The tractable subcases (bounded actions, separable utility, tree structure) are too restrictive to cover real problems."

Response: These subcases characterize *when* dimension selection becomes feasible:

- **Bounded actions:** Many real decisions have few options (buy/sell/hold, accept/reject, left-/right/straight)

- **Separable utility:** Additive decomposition is common in economics and operations research

- **Tree structure:** Hierarchical dependencies appear in configuration, organizational decisions, and causal models

The dichotomy theorem (Theorem 4.1) precisely identifies the boundary. The contribution is not that all problems are hard, but that hardness is the *default* unless special structure exists.

B.5 Objection 5: "This just formalizes the obvious"

Objection: "Everyone knows feature selection is hard. This paper just adds mathematical notation to folklore."

Response: The contribution is unification. Prior work established hardness for specific domains (feature selection in ML [11], factor identification in economics, variable selection in statistics). We prove a *universal* result that applies to *any* decision problem with coordinate structure.

This universality explains why the same "over-modeling" pattern appears across unrelated domains. It's not that each domain independently discovered the same heuristic—it's that each domain independently hit the same computational barrier.

The theorem makes "obvious" precise and proves it applies universally. This is the value of formalization.

B.6 Objection 6: "The Lean proofs don't capture the real complexity"

Objection: "The Lean formalization models an idealized version of the problem. Real coNP-completeness proofs are about Turing machines, not Lean types."

Response: The Lean formalization captures the mathematical structure of the reduction, not the Turing machine details. We prove:

1. The sufficiency-checking problem is in coNP (verifiable counterexample)

2. TAUT reduces to sufficiency checking (polynomial-time construction)

3. The reduction preserves yes/no answers (correctness)

These are the mathematical claims that establish coNP-completeness. The Turing machine encoding is implicit in Lean's computational semantics. The formalization provides machine-checked verification that the reduction is correct.

B.7 Objection 7: "The dichotomy is not tight"

Objection: "The dichotomy between $O(\log n)$ and $\Omega(n)$ minimal sufficient sets leaves a gap. What about $O(\sqrt{n})$?"

Response: The dichotomy is tight under standard complexity assumptions. The gap corresponds to problems reducible to a polynomial number of SAT instances—exactly the problems in the polynomial hierarchy between P and coNP.

In practice, the dichotomy captures the relevant cases: either the problem has logarithmic dimension (tractable) or linear dimension (intractable). Intermediate cases exist theoretically but are rare in practice.

B.8 Objection 8: "This doesn't help practitioners"

Objection: "Proving hardness doesn't help engineers solve their problems. This paper offers no constructive guidance."

   Response: Understanding limits is constructive. The paper provides:

1. **Tractable subcases** (Theorem 5.1): Check if your problem has bounded actions, separable utility, or tree structure

2. **Justification for heuristics**: Over-modeling is not laziness—it's computationally rational

3. **Focus for optimization**: Don't waste effort on optimal dimension selection; invest in good defaults and local search

   Knowing that optimal selection is coNP-complete frees practitioners to use heuristics without guilt. This is actionable guidance.

B.9 Objection 9: "But simple tools are easier to learn"

Objection: "The Simplicity Tax analysis ignores learning costs. Simple tools have lower barrier to entry, which matters for team adoption."

   Response: This objection conflates H_{central} (learning cost) with total cost. Yes, simple tools have lower learning cost. But for n use sites, the total cost is:

$$H_{\text{total}} = H_{\text{central}} + n \times H_{\text{distributed}}$$

   The learning cost is paid once; the per-site cost is paid n times. For $n > H_{\text{central}}/H_{\text{distributed}}$, the "complex" tool with higher learning cost has lower total cost.

   The objection is actually an argument for training, not for tool avoidance. If the learning cost is the barrier, pay it—the amortization makes it worthwhile.

B.10 Objection 10: "My team doesn't know metaclasses/types/frameworks"

Objection: "In practice, teams use what they know. Advocating for 'complex' tools ignores organizational reality."

   Response: The Simplicity Tax is paid regardless of whether your team recognizes it. Ignorance of the tax does not exempt you from paying it.

   If your team writes boilerplate at 50 locations because they don't know metaclasses, they pay the tax—in time, bugs, and maintenance. The tax is real whether it appears on a ledger or not.

   Organizational reality is a constraint on *implementation*, not on *what is optimal*. The Simplicity Tax Theorem tells you the optimal; your job is to approach it within organizational constraints. "We don't know X" is a gap to close, not a virtue to preserve.

B.11 Objection 11: "We can always refactor later"

Objection: "Start simple, refactor when needed. Technical debt is manageable."

   Response: Refactoring from distributed to centralized is $O(n)$ work—you are paying the accumulated Simplicity Tax all at once. If you have n sites each paying tax k, refactoring costs at least nk effort.

   "Refactor later" is not free. It is deferred payment with interest. The Simplicity Tax accrues whether you pay it incrementally (per-site workarounds) or in bulk (refactoring).

Moreover, distributed implementations create dependencies. Each workaround becomes a local assumption that must be preserved during refactoring. The refactoring cost is often *superlinear* in n.

B.12 Objection 12: "The Simplicity Tax assumes all axes are equally important"

Objection: "Real problems have axes of varying importance. A tool that covers the important axes might be good enough."

Response: The theorem is conservative: it counts axes uniformly. Weighted versions strengthen the result.

If axis a has importance w_a, define weighted tax:

$$\text{WeightedTax}(T, P) = \sum_{a \in R(P) \setminus A(T)} w_a$$

Now the incomplete tool pays $\sum w_a \times n$ while the complete tool pays 0. The qualitative result is unchanged: incomplete tools pay per-site; complete tools do not.

The "cover important axes" heuristic only works if you *correctly identify* which axes are important. By Theorem 3.6, this identification is coNP-complete. You are back to the original hardness result.

References

[1] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[2] Edoardo Amaldi and Viggo Kann. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, 209(1-2):237–260, 1998.

[3] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[4] Avrim L. Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2):245–271, 1997.

[5] David Maxwell Chickering, David Heckerman, and Christopher Meek. Large-sample learning of Bayesian networks is NP-hard. In *Journal of Machine Learning Research*, volume 5, pages 1287–1330, 2004.

[6] Stephen A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pages 151–158. ACM, 1971.

[7] Stephen A. Cook. The P versus NP problem, 2018. Millennium Prize Problems, Clay Mathematics Institute.

[8] Leonardo de Moura and Sebastian Ullrich. The lean 4 theorem prover and programming language. In *International Conference on Automated Deduction*, pages 625–635. Springer, 2021.

[9] Ronald A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A*, 222:309–368, 1922.

[10] Peter D. Grünwald. *The Minimum Description Length Principle*. MIT Press, 2007.

[11] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[12] Ronald A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, 1966.

[13] Russell Impagliazzo and Ramamohan Paturi. On the complexity of k-sat. *Journal of Computer and System Sciences*, 62(2):367–375, 2001.

[14] Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103. Springer, 1972.

[15] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[16] Erich L. Lehmann and Henry Scheffé. Completeness, similar regions, and unbiased estimation. *Sankhyā: The Indian Journal of Statistics*, 10(4):305–340, 1950.

[17] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 3rd edition, 2008.

[18] Tobias Nipkow and Gerwin Klein. *Concrete Semantics with Isabelle/HOL*. Springer, 2014.

[19] Christos H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[20] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[21] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition, 2009.

[22] Howard Raiffa and Robert Schlaifer. *Applied Statistical Decision Theory*. Harvard University Press, 1961.

[23] Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[24] Leonard J. Savage. *The Foundations of Statistics*. John Wiley & Sons, 1954.

[25] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[26] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. *Proceedings of the National Conference on Artificial Intelligence*, 21(2):1219–1226, 2006.

[27] Ilya M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[28] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, 2nd edition, 2000.

[29] Larry J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3(1):1–22, 1976.

[30] Mervyn Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–133, 1974.

[31] Marc Teyssier and Daphne Koller. Ordering-based search: A simple and effective algorithm for learning Bayesian networks. *arXiv preprint arXiv:1207.1429*, 2012.