Semantic Identity Drift in Decision Systems under Formal Certification

A Formal Separation between Operator-Based Drift and Predicate-Based Correctness

Marko A. E. Chalupa
SnapOS Foundation
audit@snapos.org

Distribution copy available via corpXiv. Canonical archived version: Zenodo DOI 10.5281/zenodo.18115847.

Abstract

Modern decision systems may satisfy formal correctness criteria while nonetheless undergoing a gradual transformation of their internal semantic identity. This paper isolates and formalizes this phenomenon as semantic identity drift.

We distinguish between two orthogonal aspects: (i) epistemic certifiability, expressed by the binary predicate $Z_{\mathrm{cert}}(L, M)$, and (ii) semantic drift, described by the time-dependent operator $D(t) : L \to L_t$ acting on problem structures. For a given pair (L, M), it may hold that $Z_{\mathrm{cert}}(L, M) = 1$, indicating the existence of valid witnesses within the model system. Independently, the semantic structure L may undergo nontrivial transformation under $D(t)$, which is not addressed by output correctness or by certification predicates.

We argue that formal correctness, optimization objectives, and telemetry are insufficient to constrain semantic identity drift, because drift operates at the level of internal meaning evolution rather than observable behavior. Accordingly, semantic stability is not entailed by epistemic success conditions. The paper therefore establishes a strict separation between operator-based drift analysis and predicate-based certification, and motivates audit methodologies that explicitly address semantic identity preservation over time.

1 Introduction

Formal methods for decision systems typically assess correctness by relating observable behavior to a specification or certification predicate. Such approaches are effective at establishing that a system satisfies given constraints at a particular point in time. However, they leave open a structurally different question: whether the internal semantic identity of a decision is preserved as the system evolves.

This paper serves as a formal separation result and makes no operational claims.

In particular, no assumptions are made about the observability, metric structure, or signal-based measurement of semantic drift. This paper addresses this question by isolating semantic identity drift as a phenomenon that is formally independent of observable behavior and output correctness. A decision system may continue to produce compliant, replayable, and formally certified outputs while the internal meaning structures that justify those decisions change over time. When this occurs, the system can no longer be regarded as making the same decision in a semantic sense, even if its behavior remains indistinguishable at the interface level.

DOI (archived version): 10.5281/zenodo.*XXXXXXX*

Preprint. Under review.

2 Preliminaries and Scope

This section fixes terminology, notational conventions, and scope assumptions used throughout the paper. Its purpose is to prevent category errors and to make explicit which distinctions are taken as primitive.

2.1 Decision Systems and Decisions

We consider a decision system as any system that, given an input context, produces an output that is intended to count as a decision relative to some specification. The term *decision* is used extensionally to refer to such outputs, but the paper is concerned with their semantic identity rather than with their execution or outcomes.

No assumptions are made about the internal realization of the system. The framework applies equally to symbolic systems, statistical models, hybrid architectures, and other formalisms, provided that an internal semantic structure can be represented abstractly.

2.2 Semantic Structures

Let L denote a semantic structure associated with a decision system at a given time. L represents the internal meaning context under which decisions are justified. The internal structure of L is left abstract; the paper does not assume a specific semantic formalism, ontology language, or representation scheme.

Semantic structures are treated as objects that may evolve over time. Temporal evolution is modeled by an operator $D(t)$, which maps an initial structure L to a time-indexed structure L_t. No assumptions are made about continuity, reversibility, monotonicity, or observability of this evolution.

When distances between semantic structures are referenced, they are assumed to be given by an explicitly declared metric or pseudometric, whose role is solely to quantify semantic separation.

2.3 Certification Predicates

Certification is modeled by a binary predicate $Z_{\mathrm{cert}}(L, M)$, defined over a semantic structure L and a model or decision system M. The predicate expresses whether valid witnesses exist that establish correctness, compliance, or derivability of M relative to L.

Certification is epistemic in nature: it concerns justification or verification relative to a fixed semantic state. The paper does not assume that certification predicates are complete, persistent over time, or preserved under semantic evolution. Certification is always understood as scoped to a specific semantic structure.

2.4 Scope of the Results

The results in this paper are purely formal. They concern logical and structural relations between semantic drift operators and certification predicates. In particular, the paper establishes non-implication results showing that semantic change and certification success are independent notions.

The paper deliberately excludes:

- empirical evaluation or benchmarks,
- detection or measurement methods for drift,
- architectural or system-level implementations,
- normative claims about system quality, governance, or responsibility.

The contribution is limited to clarifying the conceptual and formal landscape in which semantic identity, drift, and certification are discussed. Any applied interpretation of these distinctions lies outside the scope of the present work.

3 Decision Identity

This section defines decision identity and distinguishes it from related notions such as execution, correctness, and formal stability. The purpose is to fix the object of analysis before introducing semantic drift.

3.1 Definition

Decision identity is the property of a decision that specifies whether its semantic meaning, justificatory basis, and intended scope are preserved under temporal or contextual variation, independently of replayability, compliance, or formal stability.

Equivalently, decision identity denotes whether a decision remains the same decision over time, rather than merely producing the same outcome.

This definition is intentionally non-behavioral. Decision identity is not determined by output traces alone and is not reducible to execution equivalence.

3.2 Identity versus Execution

Execution concerns whether a decision procedure can be replayed to yield the same output under identical inputs. Replayability is an operational property and does not, by itself, constrain the semantic conditions under which the decision is justified.

A system may be replayable while no longer instantiating the same decision, if the semantic structure that legitimates the output has changed.

Thus, execution equivalence does not imply preservation of decision identity.

3.3 Identity versus Formal Stability

Formal stability concerns whether a system continues to satisfy a given specification, invariant, or certification predicate. Stability is therefore relative to a fixed semantic structure.

Decision identity is orthogonal to formal stability. A decision may remain formally stable while its semantic identity changes, and conversely, a decision may retain its semantic identity while formal properties fail.

Formally, formal stability is neither a sufficient nor a necessary condition for decision identity.

3.4 Identity and Semantic Drift

Semantic drift occurs at the level of the semantic structures that ground decisions. When such structures evolve, decision identity may be disrupted even in the absence of observable behavioral change.

A system may be replayable, compliant, and in equilibrium, while representing a different decision, if the reference terms, justificatory relations, or priority weights underlying the decision have changed.

This observation motivates treating decision identity as a first-class semantic property rather than as a derivative of correctness or performance.

3.5 Scope Clarification

This section introduces decision identity as a descriptive and structural notion. It does not propose criteria for detecting identity loss, nor does it assign normative weight to identity preservation.

The role of decision identity in this paper is to provide a precise target for subsequent analysis of semantic drift and non-implication results.

4 Operator-Based Semantic Drift

This section introduces semantic drift as an operator-level phenomenon acting on semantic structures. The aim is to characterize drift independently of certification outcomes, observable behavior, or performance measures.

4.1 Drift as a Structural Operator

Let L denote a semantic structure associated with a decision system at a given time. Semantic drift is modeled by a time-dependent operator

$$D(t) : L \to L_t.$$

The operator $D(t)$ is taken as primitive. No assumptions are made about its invertibility, continuity, monotonicity, or observability. In particular, no signal map or telemetry process is assumed at this level.Drift is therefore not defined by its effects on outputs, but by the possibility that the underlying semantic structure changes over time.

4.2 Drift Without Behavioral Change

Operator-based semantic drift does not presuppose any change in observable behavior. It is possible for $D(t)$ to induce a nontrivial transformation of L while the system's outputs remain invariant with respect to all tested inputs.

Accordingly, semantic drift is not defined as a deviation in performance, accuracy, or output distribution. Drift concerns internal meaning structures and may occur even when the system is replayable, compliant, and formally stable.

4.3 Drift and Decision Identity

Because semantic drift acts on the structures that ground decisions, it is directly relevant to decision identity. When $D(t)$ maps L to a structure L_t that no longer preserves the semantic conditions under which a decision was justified, decision identity may be disrupted.

This disruption is not itself a behavioral event. It is a structural change in the semantic context that legitimates decisions. As a result, decision identity may be affected by drift even when no observable deviation is present.

4.4 Drift Independent of Certification

Semantic drift is defined independently of certification predicates. The existence or absence of drift does not depend on whether $Z_{\mathrm{cert}}(L, M)$ holds at any given time. A system may satisfy certification predicates at multiple time points while undergoing semantic drift between those points.

Conversely, the failure of certification does not imply semantic drift. Certification predicates express epistemic success relative to a fixed semantic structure, whereas drift characterizes structural evolution of that structure over time.

4.5 Scope Clarification

This section treats semantic drift exclusively as an operator-based notion. It does not introduce metrics, detection criteria, or thresholds for drift, nor does it propose methods for monitoring or constraining semantic evolution.

Semantic drift, as defined here, is a structural concept. Any empirical, architectural, or normative interpretation of drift lies outside the scope of the present work.

5 Predicate-Based Certification

This section formalizes certification as a predicate-based notion and contrasts it with operator-based semantic drift. Certification is treated as an epistemic statement relative to a fixed semantic structure, not as a claim about semantic evolution.

5.1 Certification as an Epistemic Predicate

Let L denote a semantic structure and M a model or decision system. Certification is represented by a binary predicate

$$Z_{\mathrm{cert}}(L, M) \in \{0, 1\},$$

which expresses whether there exist witnesses deemed valid that establish correctness, compliance, or derivability of M relative to L.

The predicate $Z_{\mathrm{cert}}(L, M) = 1$ is an epistemic statement. It asserts the existence of sufficient justification under the semantic assumptions encoded in L. No claim is made about completeness, optimality, or persistence of such justification beyond the scope of the given semantic structure.

5.2 Scope and Fixity of Certification

Certification predicates are always scoped to a specific semantic structure. The truth value of $Z_{\mathrm{cert}}(L, M)$ is evaluated relative to L and does not, by itself, extend to other semantic structures that may arise through semantic evolution.

If the semantic structure changes from L to L_t, the predicate $Z_{\mathrm{cert}}(L, M)$ does not entail any corresponding statement about $Z_{\mathrm{cert}}(L_t, M)$. Re-certification is a separate epistemic act and is not implied by prior certification success.

5.3 Certification Without Semantic Stability

Certification predicates do not constrain semantic stability. A system may satisfy

$$Z_{\mathrm{cert}}(L, M) = 1$$

at multiple time points while the underlying semantic structure undergoes nontrivial evolution between those points.

Certification success therefore does not imply preservation of semantic identity. Certification establishes epistemic success relative to a given semantic state but places no constraints on how that state may evolve over time.

5.4 Failure of Certification and Drift

Conversely, failure of certification does not imply semantic drift. The predicate

$$Z_{\mathrm{cert}}(L, M) = 0$$

may result from insufficient evidence, incomplete specifications, or limitations of the certification procedure, even when the semantic structure L remains unchanged.

Certification predicates thus provide no direct information about whether semantic drift has occurred. They express epistemic success or failure under given assumptions, not structural change.

5.5 Scope Clarification

This section treats certification exclusively as a predicate-based notion. It does not address how certification is obtained, how witnesses are constructed, or how certification procedures are implemented.

The role of predicate-based certification in this paper is to serve as a formal counterpart to operator-based semantic drift. Together, these notions support the non-implication results established in the subsequent section.

6 Non-Implication Results

This section states the formal separation results that follow from the preceding definitions. The results establish that operator-based semantic drift and predicate-based certification are logically independent notions. No constructive procedures or detection criteria are assumed.

All propositions follow directly from the definitions by exhibiting admissible models in which semantic structure, observation, and certification are formally decoupled.

6.1 No Entailment from Observational Equivalence

Proposition 1 (No Entailment from Observational Equivalence)**.** *Observational equivalence of system behavior does not entail stability of internal semantic structure.*

Assume an observational equivalence relation $\sim$ on system executions that abstracts from internal semantic structure. Then there exist a semantic structure L, a drift operator $D(t)$, and a system M such that:

- the executions of M remain observationally equivalent over time with respect to $\sim$, and
- $D(t)$ induces a nontrivial transformation $L \to L_t$.

In such cases, semantic drift may occur without any corresponding observable deviation. Observational equivalence therefore provides no guarantee of semantic identity preservation.

Proof sketch. Choose $D(t)$ to act nontrivially on semantic components that are not referenced by the observation relation $\sim$. By construction, observational equivalence is preserved while the internal semantic structure evolves. □

6.2 Certification Does Not Imply Absence of Drift

Proposition 2 (No Certification Entailment)**.** *Certification success does not imply the absence of semantic drift.*

There exist semantic structures L, systems M, and times $t_1 < t_2$ such that

$$Z_{\mathrm{cert}}(L, M) = 1 \quad \text{and} \quad Z_{\mathrm{cert}}(L_{t_2}, M) = 1,$$

while the drift operator $D(t)$ induces a semantic transformation between L and L_{t_2} that disrupts decision identity.

Certification predicates place no constraints on the transformation path induced by $D(t)$ between t_1 and t_2. Certification is evaluated pointwise with respect to a given semantic structure and does not constrain how that structure evolves over time.

Proof sketch. Let Z_{cert} depend only on semantic components that are preserved at the certification points. Choose $D(t)$ to modify identity-relevant components between t_1 and t_2 while leaving the certification-relevant components unchanged at those points. □

6.3 Absence of Certification Does Not Imply Drift

Proposition 3 (No Drift from Certification Failure)**.** *Failure of certification does not imply semantic drift.*

There exist cases in which

$$Z_{\mathrm{cert}}(L, M) = 0,$$

while the semantic structure L remains unchanged. Certification failure may result from epistemic limitations, incomplete specifications, or unavailable witnesses, without any structural semantic change.

Proof sketch. Hold the semantic structure L fixed and vary only the epistemic conditions required for witness construction or verification. Certification failure follows without invoking any semantic evolution. □

6.4 Mutual Independence

Taken together, Propositions 6.1–6.3 establish a non-implication relationship between operator-based semantic drift and predicate-based certification:

- Semantic drift neither entails nor is entailed by observational or behavioral stability.

- Certification success neither entails nor is entailed by semantic stability.
- Certification failure neither entails nor is entailed by semantic drift.

These statements are separation results. They assert the absence of logical entailment, not the presence of correlation, causal interaction, or empirical likelihood.

6.5 Consequence for Analysis

The non-implication results show that semantic identity is not characterized by correctness predicates or observable behavior alone. Any analysis that conflates certification outcomes with semantic stability treats epistemic predicates as constraints on structural evolution.

The results therefore justify treating operator-based semantic drift and predicate-based certification as distinct analytical dimensions. Subsequent sections rely on this separation without introducing additional assumptions or evaluative claims.

7 Audit Interpretation

This section interprets the formal separation results established in Section 6 from an audit perspective. It introduces no new assumptions and does not extend the formal framework. Its purpose is to clarify what audits can and cannot infer given the logical independence between operator-based semantic drift and predicate-based certification.

7.1 Audit as Structural Analysis

In this paper, an audit is understood as an analysis that inspects relationships between semantic structures across time. An audit does not operate on outputs alone and does not coincide with certification. Instead, it concerns whether semantic identity is preserved as the underlying semantic structure evolves.

Under this interpretation, audits are not decision procedures and do not yield binary success or failure outcomes. They are comparative analyses over semantic states, evaluated relative to transformation paths rather than single states.

7.2 Limits of Certification-Based Audits

The non-implication results show that certification predicates do not provide information about semantic identity preservation. Certification establishes epistemic success at a given semantic state but does not constrain how that state may evolve.

As a consequence, an audit that relies solely on certification outcomes cannot determine whether semantic identity has been preserved between certification points. Certification success at multiple times does not entail identity preservation, and certification failure does not entail semantic change.

This limitation is structural rather than methodological. It follows directly from the predicate-based nature of certification and is independent of how witnesses are constructed or evaluated.

7.3 Audit Targets and Objects of Comparison

Given the formal separation established in Sections 4–6, the proper target of audit analysis is the evolution of semantic structures under the drift operator. Audits compare semantic states L and L_t with respect to identity-relevant components, rather than comparing outputs, metrics, or certification results.

Such comparisons are path-sensitive: they depend on how semantic structures change over time, not merely on their endpoints. This distinguishes audit interpretation from pointwise evaluation.

7.4 Audit Without Detection Claims

Nothing in the formal results implies that semantic drift is detectable, measurable, or observable in general. The audit interpretation developed here does not assert that drift can be identified from data, logs, or system outputs, nor that it can be algorithmically detected.

The role of audit interpretation in this paper is purely conceptual. It specifies which questions an audit can meaningfully address, given the logical independence between drift and certification, and which questions lie outside the reach of predicate-based analysis.

This interpretation does not presuppose the existence of telemetry or signal processes; it remains valid even when no measurable observables are available.

7.5 Scope Clarification

This section does not propose audit procedures, tools, standards, or architectures. It does not assign normative force to audits, nor does it link audit interpretation to governance, compliance, or responsibility frameworks.

The contribution is limited to clarifying the interpretive role of audits in light of the formal separation results: audits concern semantic structure over time, whereas certification concerns epistemic status at a fixed semantic state. Any operationalization of this distinction lies beyond the scope of the present work.

8 Discussion

This paper has deliberately restricted its scope to formal distinctions and non-implication results. The discussion therefore does not extend the theory, propose applications, or introduce new assumptions. Instead, it clarifies the implications and limits of the results established in Sections 3–7.

8.1 What the Results Do and Do Not Establish

The core contribution of the paper is a separation: operator-based semantic drift and predicate-based certification are formally independent notions. This separation is negative in character. It does not assert that semantic drift necessarily occurs, nor that certification is ineffective or misguided. Rather, it establishes that neither behavioral stability nor certification success suffices to characterize semantic identity preservation.

Accordingly, the results should not be read as claims about prevalence, risk, or severity of semantic drift in real systems. They identify a logical possibility space, not an empirical diagnosis.

8.2 Implications for Formal Analysis

From a formal perspective, the results highlight a category distinction that is often blurred in system analysis. Certification predicates are epistemic and pointwise: they assess derivability or correctness relative to a fixed semantic structure. Drift operators are structural and temporal: they describe how semantic structures may evolve over time.

Treating these notions as interchangeable leads to invalid inferences. This is not a methodological oversight but a logical consequence of the underlying definitions. Any framework that collapses predicates and operators into a single evaluative dimension risks conflating epistemic success with structural stability.

8.3 Relation to Existing Formal Traditions

The separation articulated in this paper aligns with established distinctions in formal semantics and system theory, where internal semantic state is treated independently of observable behavior or specification satisfaction. The present work does not extend these traditions but applies their core insight to the context of decision identity and long-horizon system evolution.

Importantly, the paper does not rely on any specific semantic formalism. The discussion therefore remains agnostic with respect to representation choices, logics, or modeling frameworks.

8.4 Limits of the Present Work

Several limitations follow directly from the paper's design. No criteria are given for detecting, measuring, or constraining semantic drift. No assumptions are made about the likelihood of drift in deployed systems. No guidance is offered for system design, governance, or evaluation.

These omissions are intentional. Addressing them would require additional assumptions—empirical, architectural, or normative—that lie outside the scope of a purely formal separation result.

8.5 Interpretive Caution

Finally, the discussion emphasizes that the results do not license general conclusions about system reliability, trustworthiness, or responsibility. The formal independence of drift and certification constrains what can be inferred, not what ought to be done.

The value of the results lies in preventing category errors. They clarify which inferences are invalid even in the presence of correct behavior and successful certification. Any further interpretation or application must respect this boundary.

9 Conclusion

This paper has presented a formal separation between operator-based semantic drift and predicate-based certification in decision systems. By modeling semantic drift as a transformation operator on semantic structures and certification as an epistemic predicate evaluated relative to a fixed structure, the paper makes explicit that these notions are logically independent.

The central results are negative in character. They show that neither observable behavioral stability nor certification success entails preservation of semantic identity, and that certification failure does not imply semantic drift. These non-implication results follow directly from the definitions and do not rely on empirical assumptions, architectural choices, or normative criteria.

By clarifying this separation, the paper identifies a precise limitation of correctness- and certification-based analysis. Such predicates characterize epistemic status at a given semantic state but do not constrain how that state may evolve over time. Semantic identity, when treated as a structural property, therefore lies outside the scope of pointwise correctness predicates.

The contribution of the paper is intentionally narrow. It does not propose mechanisms for detecting drift, methods for enforcing semantic stability, or prescriptions for system design or governance. Instead, it provides a formal baseline that delineates what can and cannot be inferred about semantic identity from certification outcomes and observable behavior. This baseline establishes the minimal conceptual ground on which further analysis can proceed without category error.

Subsequent work may introduce metrics, signals, or telemetry to study detectability of semantic drift, but such assumptions lie strictly beyond the scope of the present separation result.

References

[1] M. A. E. Chalupa. *Semantic Identity Drift in Decision Systems under Formal Certification*. Zenodo preprint, 2025. DOI: 10.5281/zenodo.18115847.